

Citation for published version:

Shaw, SC & Goldstein, M 2012, 'Finite population corrections for multivariate Bayes sampling', *Journal of Statistical Planning and Inference*, vol. 142, no. 10, pp. 2844-2861. <https://doi.org/10.1016/j.jspi.2012.04.001>

DOI:

[10.1016/j.jspi.2012.04.001](https://doi.org/10.1016/j.jspi.2012.04.001)

Publication date:

2012

Document Version

Peer reviewed version

[Link to publication](#)

NOTICE: this is the author's version of a work that was accepted for publication in Journal of Statistical Planning and Inference. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Journal of Statistical Planning and Inference, 2012, vol 142, issue 10, DOI# /10.1016/j.jspi.2012.04.001

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Finite population corrections for multivariate Bayes sampling

By Simon C. Shaw

Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, U.K.

s.c.shaw@maths.bath.ac.uk

and Michael Goldstein

*Department of Mathematical Sciences, University of Durham, Science Laboratories,
South Road, Durham, DH1 3LE, U.K.*

michael.goldstein@durham.ac.uk

ABSTRACT

We consider the adjustment, based upon a sample of size n , of collections of vectors drawn from either an infinite or finite population. The vectors may be judged to be either Normally distributed or, more generally, second-order exchangeable. We develop the work of Goldstein and Wooff (1998) to show how the familiar univariate finite population corrections (fpc) naturally generalise to individual quantities in the multivariate population. The types of information we gain by sampling are identified with the orthogonal canonical variable directions derived from a generalised eigenvalue problem. These canonical directions share the same co-ordinate representation for all sample sizes and, for equally defined individuals, all population sizes enabling simple comparisons between both the effects of different sample sizes and of different population sizes. We conclude by considering how the fpc is modified for multivariate cluster sampling with exchangeable clusters. In univariate two-stage cluster sampling we may decompose the variance of the population mean into the sum of the variance of cluster means and the variance of the cluster members within clusters. The first term has a fpc relating to the sampling fraction of clusters, the second term has a fpc relating to the sampling fraction of cluster size. We illustrate how this generalises in the multivariate case. We decompose the variance into two terms: the first relating to multivariate finite population sampling of clusters and the second to multivariate finite population sampling within clusters. We solve two generalised eigenvalue problems to show how to generalise the univariate to the multivariate: each of the two fpcs attaches to one, and only one, of the two eigenbases.

Keywords: finite population correction; two-stage cluster sampling; canonical directions; canonical resolutions; second-order exchangeability; Bayes linear methods.

1 Introduction

A fundamental result in sampling theory is the finite population correction (fpc) formula for a random sample without replacement from a finite population. The fpc corrects the

variance of a sample mean, to take account of the sampling fraction for the data. This allows us to judge how large the population must be in order to ignore such corrections, and gives a simple multiplier to reduce the variance when this correction is not ignorable.

The Bayesian counterpart of the fpc is of similar form. The direct counterpart to the sample theoretic fpc is the corresponding Bayes linear mean and variance for a population mean based on a random sample without replacement from a finite population whose members are judged exchangeable, as this analysis is also based only on mean and variance judgements. The full Bayes counterpart to this fpc corresponds to the Gaussian likelihood and prior distribution for the population mean of the finite collection. For either version, the fpc again corrects the variance for the sample mean, providing a simple determination of sufficient sample size for finite population Bayesian sampling.

However, each fpc above relates to a univariate sample. When we take a sample of exchangeable vectors from a finite population, the corresponding form for the fpc depends on a matrix inverse which requires evaluation for each choice of sample size. It is therefore not straightforward to determine whether the sampling fraction is ignorable when planning a multivariate Bayesian sampling design for a finite population and, when the correction is not ignorable, to provide a simple guide to sample size determination. The problem is further complicated by the requirement that we may also wish to learn about linear combinations of elements of the population mean vector.

In this paper, we show that there is a simple and natural representation of the fpc for multivariate Bayesian sampling from a finite population, which is suitable for planning multivariate sampling designs, illustrating the use of the multivariate fpc for setting appropriate sample sizes. We then show that the results carry over to the case of two-stage multivariate cluster sampling where each stage can be viewed as sampling exchangeable vectors from a finite population. The paper proceeds as follows. In §2 we recall the fpc in classical univariate simple random sampling and compare this to the Bayesian analogue using normal modelling both in the univariate and multivariate case. In §3 we show how the univariate fpc generalises to individual quantities in the multivariate population and show how we can easily compare both different sample sizes and different population sizes. In §4 we illustrate the theory using an example concerning examination data. We extend the model to encompass two-stage cluster sampling in §5, concluding in §6 with an example illustrating the theory.

2 Univariate and multivariate sampling

2.1 Classical univariate simple random sampling

We draw a sample of n individuals from a population and make a single measurement, X , on each individual. We let X_i denote the measurement for the i th individual and $\bar{X} = (1/n) \sum_{i=1}^n X_i$ the sample mean. Suppose that the population is infinite with mean μ and variance $\sigma^2 = 1/r$ and that we take a simple random sample. Then,

$$E(\bar{X} | \mu, \sigma) = \mu; \text{Var}^{-1}(\bar{X} | \mu, \sigma) = nr. \quad (1)$$

Consider, instead, the scenario when the population is finite, containing only N members. We make the following definition.

Definition 1 *The finite population correction (fpc) for a sample of size n drawn from a population of size N is defined to be*

$$g(N, n) = \left(1 - \frac{n-1}{N-1}\right)^{-1}. \quad (2)$$

Let the population mean be $\mu_N = (1/N) \sum_{i=1}^N X_i$ and the population variance $\sigma_N^2 = (1/N) \sum_{i=1}^N (X_i - \mu_N)^2 = 1/r_N$. For a random sample, without replacement, we find that

$$E(\bar{X} | \mu_N, \sigma_N) = \mu_N; \text{Var}^{-1}(\bar{X} | \mu_N, \sigma_N) = g(N, n)nr_N. \quad (3)$$

Notice how the fpc, as given by (2), expresses the increase in sampling precision as a fraction of the population observed.

2.2 Bayesian univariate sampling theory

The classical model of simple random sampling may be viewed, in a Bayesian context, as a judgement of exchangeability. If the infinite sequence of random quantities X_1, X_2, \dots are judged exchangeable then, conditional on an unknown distribution function F , X_1, X_2, \dots are independent (de Finetti, 1937; Hewitt and Savage, 1955). One conventional model is that the X_i s are judged to be independent and identically distributed as $N(\mu, 1/r)$, where the prior for μ is $N(\mu_0, 1/r_0)$. Hence, we judge our prior for F to be only non-zero on the subspace of normal distributions with precision r . The posterior for μ given X_1, \dots, X_n is $N(\mu_n, 1/r_n)$ where

$$\mu_n = \frac{r_0\mu_0 + nr\bar{X}}{r_0 + nr}; \quad r_n = r_0 + nr. \quad (4)$$

Note that if $r_0 \rightarrow 0$, to represent weak prior information about μ , then $\mu_n \rightarrow \bar{X}$ and $r_n \rightarrow nr$ corresponding exactly with the classical simple random sampling approach, see (1). The corresponding formulation when we sample from a finite population, of size N , is to view the X_i s as independent and identically distributed $N(\theta, 1/r)$ where the prior distribution of θ is $N(\mu_0, 1/r_0)$. This framework was developed by Ericson (1969) with extensions by, for example, Royall and Pfeffermann (1982), Smouse (1984) and Bolfarine (1990). O'Hagan and Forster (2004, §14.30) and Ghosh and Meeden (1997) provide textbook discussions. The prior for μ_N is thus $N(\mu_0, 1/r_{(N;0)})$ where $r_{(N;0)}^{-1} = r_0^{-1} + (1/N)r^{-1}$ and the posterior, given X_1, \dots, X_n , for μ_N is $N(\mu_{(N;n)}, 1/r_{(N;n)})$ where

$$\mu_{(N;n)} = \frac{r_{(N;0)}\mu_0 + g(N, n)nr_{(N)}\bar{X}}{r_{(N;0)} + g(N, n)nr_{(N)}}; \quad r_{(N;n)} = r_{(N;0)} + g(N, n)nr_{(N)}, \quad (5)$$

and $r_{(N)} = \text{Var}^{-1}(X_i | \mu_N) = (1 - \frac{1}{N})^{-1}r$. In this Bayesian setting (5) relates to (4) in exactly the same way as (3) relates to (1) in the classical framework.

2.3 Bayesian multivariate sampling theory

We now explore whether the simplicity of the univariate approach, exhibited by the posterior means and precisions given in (4) and (5), remains when we make multivariate measurements. Suppose that we wish to make the same series of measurements $\mathcal{C} = \{X_1, \dots, X_{v_0}\}$ on each individual in a sample. The measurements for the i th such individual are denoted by $\mathcal{C}_i = \{X_{1i}, \dots, X_{v_0i}\}$. The full population collection is formed as the union of all of the elements in all of the collections, \mathcal{C}_i , and is denoted by \mathcal{C}^* .

Consider the case when the population is infinite and we judge the collections $\mathcal{C}_1, \mathcal{C}_2, \dots$ to be independent and identically distributed multivariate normal random quantities with expectation vector $\mu(\mathcal{C}) = [\mu(X_1) \dots \mu(X_{v_0})]^T$ and known precision matrix R . The prior

for $\mu(\mathcal{C})$ is also judged to be multivariate normal with known expectation vector m_0 and known precision matrix R_0 . For simplicity of exposition we assume that R and R_0 are positive definite, otherwise the corresponding Moore-Penrose inverses should be used to obtain equivalent results to those that follow. We observe the measurements for the first n individuals, $\mathcal{C}(n) = \cup_{i=1}^n \mathcal{C}_i$ and let $\bar{\mathcal{C}}(n) = \{\bar{X}_1, \dots, \bar{X}_{v_0}\}$ denote the collection of sample averages. Given $\mathcal{C}(n)$, the posterior distribution for $\mu(\mathcal{C})$ is multivariate normal with expectation vector m_n and precision matrix R_n where

$$m_n = \{R_0 + nR\}^{-1}\{R_0 m_0 + nR\bar{\mathcal{C}}(n)\}; \quad R_n = R_0 + nR. \quad (6)$$

The multivariate case mimics the univariate: the univariate quantities in (4) are replaced by their matrix equivalents in (6). The same generalisation occurs when the population consists of only N individuals; the total collection being $\mathcal{C}(N) = \cup_{i=1}^N \mathcal{C}_i$. Let $\mu_N(X_v) = \frac{1}{N} \sum_{i=1}^N X_{vi}$ denote the population mean for the v th measurement and $\mu_N(\mathcal{C}) = \{\mu_N(X_1), \dots, \mu_N(X_{v_0})\}$ the collection of population means. We judge that the distribution of any subset of the $\mathcal{C}(N)$ is identical to that for an equivalently sized subset of individuals drawn from the infinite population. The prior distribution of $\mu_N(\mathcal{C})$ is thus normal with expectation vector m_0 and precision matrix $R_{(N;0)}$ where $R_{(N;0)}^{-1} = R_0^{-1} + \frac{1}{N}R^{-1}$. The posterior distribution for $\mu_N(\mathcal{C})$, given $\mathcal{C}(n)$, is multivariate normal with expectation vector $m_{(N;n)}$ and precision matrix $R_{(N;n)}$ where

$$m_{(N;n)} = \{R_{(N;0)} + g(N, n)nR_{(N)}\}^{-1}\{R_{(N;0)}m_0 + g(N, n)nR_{(N)}\bar{\mathcal{C}}(n)\}; \quad (7)$$

$$R_{(N;n)} = R_{(N;0)} + g(N, n)nR_{(N)}, \quad (8)$$

and $R_{(N)} = \frac{N}{N-1}R$ is the posterior precision for \mathcal{C}_i given $\mu_N(\mathcal{C})$. We observe the mimicry between the univariate and multivariate case in the finite population setting: in (7) and (8) the matrix equivalents have replaced the corresponding univariate quantities in (5).

However, the apparent simplicity of the multivariate generalisation, displayed by (6) - (8), conceals the complex way in which changes in n modify the posterior uncertainties for the individual quantities. We now extend Goldstein and Wooff (1998) to show that there is a natural generalisation from the univariate to the multivariate case which preserves the simplicity of the relationship between sample size and posterior mean and precision for individual quantities within the population collection.

3 Finite population corrections for Bayes linear analysis

3.1 Sampling theory using Bayes linear methods

The infinite models described in §2.2 and §2.3 required a further assumption, that of normality, to that of exchangeability. For example, the finite model in §2.2 involves the use of a hyperparameter, θ , to generate the joint prior distribution of X_1, \dots, X_N . As Ericson (1969; p198) writes, ‘the generation of a joint prior distribution by this approach is, barring differences in probabilistic interpretation, equivalent to viewing the finite population as a sample from an infinite superpopulation having unknown parameter θ .’ It is implicit in this approach that the finitely exchangeable sequence X_1, \dots, X_N may be embedded in an infinitely exchangeable sequence of equivalently defined random quantities. However, see for example Bernardo and Smith (1994; p171), a finitely exchangeable sequence cannot always be embedded into a larger finitely exchangeable sequence, much less an infinitely exchangeable sequence.

An alternative approach is to judge the population members to be second-order exchangeable (Goldstein, 1986) and to update these beliefs using the Bayes linear methodology where the adjusted expectation of a random quantity X given observation of a collection of quantities D is denoted $E_D(X)$ and the corresponding adjusted variance is denoted $Var_D(X)$; see Chapter 3 of Goldstein and Wooff (2007) for a detailed explanation of Bayes linear belief adjustment and Goldstein (1999) for an overview. As Hartigan (1969; p447) points out if X and D are jointly normally distributed, then the adjusted quantities coincide with the usual definitions of conditional expectation, that is $E_D(X) = E(X | D)$ and $Var_D(X) = Var(X | D)$. By virtue of proceeding from only a second-order specification, this Bayes linear approach to population sampling under second-order exchangeability is closer, in spirit, to the classical simple random sampling approach described in §1.1.

3.2 Sampling from an infinite population

We now consider the problem of §2.3 using a model of second-order exchangeability and adjustment using the Bayes linear approach. For second-order exchangeability, as Goldstein (1986; p973) writes, ‘what is actually required is the consideration of two cases, with all other values following from the perceived “symmetries” between cases’: our considerations are not dependent upon the size of the population. In this section we shall consider the case when the full collection \mathcal{C}^* is formed from an infinite number of individuals.

Assumption 1 *We judge that the collection \mathcal{C} is second-order exchangeable over the full collection \mathcal{C}^* . For all $i \neq j$ our second-order specifications thus take the form*

$$E(\mathcal{C}_i) = m_0; Var(\mathcal{C}_i) = D; Cov(\mathcal{C}_i, \mathcal{C}_j) = C.$$

Using the second-order representation theorem of Goldstein (1986) we may write

$$\mathcal{C}_i = \mu(\mathcal{C}) + R_i(\mathcal{C}) \quad (9)$$

where $\mu(\mathcal{C})$ in the limit, in mean square, of $\mu_N(\mathcal{C}) = \frac{1}{N} \sum_{i=1}^N \mathcal{C}_i$ and $R_i(\mathcal{C}) = \mathcal{C}_i - \mu(\mathcal{C})$. For all i , the $R_i(\mathcal{C})$ are mutually uncorrelated and also uncorrelated with $\mu(\mathcal{C})$. We consider the adjustment of our beliefs following the observation of $\mathcal{C}(n) = \cup_{i=1}^n \mathcal{C}_i$. In this context, the conditional results given by (6) will match our adjusted beliefs for $\mu(\mathcal{C})$ when $D = R^{-1} + R_0^{-1}$ and $C = R_0^{-1}$ though we now consider individual quantities in the population collection.

Let $\langle \mathcal{C} \rangle$ denote the collection of linear combinations $Z = \sum_{v=1}^{v_0} \alpha_v X_v$ of elements of \mathcal{C} . For each $Z \in \langle \mathcal{C} \rangle$ and each individual i , we may construct Z_i , the value of Z for individual i , as $Z_i = \sum_{v=1}^{v_0} \alpha_v X_{vi}$. We construct linear combinations $\mu(Z) = \sum_{v=1}^{v_0} \alpha_v \mu(X_v)$ of the elements of $\mu(\mathcal{C})$ whilst linear combinations of the sample averages $\bar{\mathcal{C}}(n)$ are denoted by $\bar{Z} = \sum_{v=1}^{v_0} \alpha_v \bar{X}_v$. $\langle \mathcal{C}_i \rangle$, $\langle \mu(\mathcal{C}) \rangle$, $\langle \bar{\mathcal{C}}(n) \rangle$ respectively denote the collection of linear combinations of the elements of \mathcal{C}_i , $\mu(\mathcal{C})$ and $\bar{\mathcal{C}}(n)$. Hence, the labelling convention is such that for any $Z \in \langle \mathcal{C} \rangle$, Z_i , $\mu(Z)$ and \bar{Z} share the same coordinate representation.

Definition 2 *The underlying canonical variable directions are defined as the columns of the matrix $W = [W_1 \dots W_{v_0}]$ where $W_s = RH_s$ and H_s is the s th column of the matrix H solving the generalised eigenvalue problem*

$$RH = (R + R_0)H\Phi \quad (10)$$

where $\Phi = \text{diag}(\phi_1, \dots, \phi_{v_0})$ is the matrix of eigenvalues, $R^{-1} = D - C$ and $R_0^{-1} = C$. H is normed so that $H^T R H = I$ and $H^T (R + R_0) H \Phi = I$. The ordered eigenvalues $1 > \phi_1 \geq \dots \geq \phi_{v_0} > 0$ are termed the underlying canonical variable resolutions.

Let W_{vs} denote the v th component of the s th canonical variable direction and, for each $s = 1, \dots, v_0$, define $Y_s \in \langle \mathcal{C} \rangle$ to be $Y_s = \sum_{v=1}^{v_0} W_{vs} X_v$. The underlying canonical variable directions and resolutions provide the wherewithal to generalise the univariate results of §2.2 to individual quantities in the multivariate population. The following theorem may be obtained as a reformulation of Theorem 3 of Goldstein and Wooff (1998).

Theorem 1 *For a sample of size n drawn from an infinite population, the collection $\mu(\mathcal{Y}) = \{\mu(Y_1), \dots, \mu(Y_{v_0})\}$ form a basis for $\langle \mu(\mathcal{C}) \rangle$. The $\mu(Y_s)$ are a priori uncorrelated, and, for all samples of any size, a posteriori uncorrelated. The posterior adjusted expectation, $\mu_{ns} = E_{\mathcal{C}(n)}\{\mu(Y_s)\}$, and posterior adjusted precision, $r_{ns} = \text{Var}_{\mathcal{C}(n)}^{-1}\{\mu(Y_s)\}$ for $\mu(Y_s)$ are given by*

$$\mu_{ns} = \frac{r_{0s}\mu_{0s} + nr_s\bar{Y}_s}{r_{0s} + nr_s}; \quad r_{ns} = r_{0s} + nr_s, \quad (11)$$

where μ_{0s} , r_{0s} are the prior expectation and precision for $\mu(Y_s)$ and r_s is the adjusted precision for any individual Y_{si} given $\mu(\mathcal{C})$.

The collection of univariate quantities in (11) are expressed identically to those in (4). There is a natural generalisation from the univariate to the multivariate for individual quantities in the population collection $\langle \mu(\mathcal{C}) \rangle$ and these quantities remain the same for each choice of n . The qualitative and quantitative features of the update remain the same no matter the sample size and may be obtained from the solution of a single generalised eigenvalue problem. Note that the actual values of r_{0s} and r_s are easily obtained. Under the scalings used in Definition 2 we have that, for each $s = 1, \dots, v_0$, $r_{0s} = \phi_s^{-1}(1 - \phi_s)$ and $r_s = 1$. However, the ϕ_s have a more fundamental interpretive role. Let Res_{ns} denote the resolution of $\mu(Y_s)$ given $\mathcal{C}(n)$, that is Res_{ns} is the proportion of variance of $\mu(Y_s)$ resolved by the observation of $\mathcal{C}(n)$. We have

$$\text{Res}_{ns} = 1 - \frac{r_{0s}}{r_{ns}} = \frac{n\phi_s}{(n-1)\phi_s + 1} \quad (12)$$

so that $\phi_s = \text{Res}_{1s}$ is the resolution of $\mu(Y_s)$ given $\mathcal{C}(1)$: the proportion of variance of $\mu(Y_s)$ resolved by a single observation. As the collection $\mu(\mathcal{Y})$ form a basis for $\langle \mu(\mathcal{C}) \rangle$ then for any $Z \in \langle \mathcal{C} \rangle$ we have $\mu(Z) = \sum_{s=1}^{v_0} \text{Cov}\{\mu(Z), \mu(Y_s)\} r_{0s} \mu(Y_s)$ from which

$$\begin{aligned} E_{\mathcal{C}(n)}\{\mu(Z)\} &= \sum_{s=1}^{v_0} \text{Cov}\{\mu(Z), \mu(Y_s)\} r_{0s} \mu_{ns}; \\ \text{Var}_{\mathcal{C}(n)}\{\mu(Z)\} &= \sum_{s=1}^{v_0} \text{Cov}\{\mu(Z), \mu(Y_s)\}^2 r_{0s}^2 r_{ns}^{-1}. \end{aligned}$$

Hence, from the ordering of the ϕ_s and (12), we see that for a sample of size n , subject to being uncorrelated with $\mu(Y_1), \dots, \mu(Y_j)$, quantities proportional to $\mu(Y_{j+1})$ have the largest resolution or equivalently the smallest ratio of posterior to prior variance. As Goldstein and Wooff (1998) explain, for each choice of n , $\mu(\mathcal{Y})$ forms an orthogonal grid over $\langle \mu(\mathcal{C}) \rangle$ for which we expect to learn most, in terms of variance reduction, about those quantities with large correlations with the early $\mu(Y_s)$. It is straightforward to utilise (12), see for example Corollary 1 of Goldstein and Wooff (1998), to simplify any design problem for which it is necessary to choose the sample size required to achieve a specified variance reduction over elements of $\langle \mu(\mathcal{C}) \rangle$.

3.3 Sampling from a finite population

We now show that a similar generalisation occurs when we explore the finite case. We operate under the model given by Assumption 1, the only difference being that we now assume that the full collection \mathcal{C}^* contains N individuals rather than an infinite number. In this finite setting, from Goldstein (1986), we may write

$$\mathcal{C}_i = \mu_N(\mathcal{C}) + R_{N,i}(\mathcal{C}) \quad (13)$$

where $\mu_N(\mathcal{C}) = \frac{1}{N} \sum_{i=1}^N \mathcal{C}_i$ and $R_{N,i}(\mathcal{C}) = \mathcal{C}_i - \mu_N(\mathcal{C})$. For all i , the $R_{N,i}(\mathcal{C})$ are uncorrelated with $\mu_N(\mathcal{C})$ but the finite nature of the population induces a correlation between $R_{N,i}(\mathcal{C})$ and $R_{N,j}(\mathcal{C})$ for each $i \neq j$, the correlation being to the order of $\frac{1}{N}$. Note that there is no need for an implicit assumption of a superpopulation from which the finite population is a sample from. The justification for the prior specification and use of $\mu_N(\mathcal{C})$ comes directly from a judgement of second-order exchangeability and the resulting representation theorem, (13). The natural relationship between $\mu(\mathcal{C})$ and $\mu_N(\mathcal{C})$ also follows from this: the former is the limit, in mean square, of the latter as $N \rightarrow \infty$.

Letting $\langle \mu_N(\mathcal{C}) \rangle$ denote the collection of linear combinations $\mu_N(Z) = \sum_{v=1}^{v_0} \alpha_v \mu_N(X_v)$ of the elements of $\mu_N(\mathcal{C})$, we consider the adjusted distribution for quantities contained in $\langle \mu_N(\mathcal{C}) \rangle$ given $\mathcal{C}(n)$. We have the following theorem; the proof is in the appendix.

Theorem 2 *For a sample of size n drawn from a population of size N , the collection $\mu_N(\mathcal{Y}) = \{\mu_N(Y_1), \dots, \mu_N(Y_{v_0})\}$ forms a basis for $\langle \mu_N(\mathcal{C}) \rangle$. The $\mu_N(Y_s)$ are a priori uncorrelated, and, for all samples of any size, a posteriori uncorrelated. The posterior adjusted expectation, $\mu_{(N;n)s} = E_{\mathcal{C}(n)}\{\mu_N(Y_s)\}$, and posterior adjusted precision, $r_{(N;n)s} = \text{Var}_{\mathcal{C}(n)}^{-1}\{\mu_N(Y_s)\}$, for $\mu_N(Y_s)$ are given by*

$$\mu_{(N;n)s} = \frac{r_{(N;0)s}\mu_{0s} + g(N, n)nr_{(N)s}\bar{Y}_s}{r_{(N;0)s} + g(N, n)nr_{(N)s}}; \quad (14)$$

$$r_{(N;n)s} = r_{(N;0)s} + g(N, n)nr_{(N)s}, \quad (15)$$

where μ_{0s} , $r_{(N;0)s}$ are the prior expectation and precision for $\mu_N(Y_s)$, $r_{(N)s}$ is the adjusted precision for Y_{si} given $\mu_N(\mathcal{C})$, and $g(N, n)$ the fpc.

The univariate quantities in (14) and (15) are expressed identically to those in (5) showing that there is also a natural generalisation for quantities in the population collection $\langle \mu_N(\mathcal{C}) \rangle$ in this finite setting as well as in the infinite setting. Once again, these quantities remain the same for each choice of n and can be derived from the solution of the generalised eigenvalue problem given in Definition 2. The quantitative information may also be directly obtained from this solution. Letting $Res_{(N;n)s}$ denote the resolution of $\mu_N(Y_s)$ given $\mathcal{C}(n)$ then, from the proof to Theorem 2, we have that

$$Res_{(N;n)s} = 1 - \frac{r_{(N;0)s}}{r_{(N;n)s}} = \frac{n\{(N-1)\phi_s + 1\}}{N\{(n-1)\phi_s + 1\}} \quad (16)$$

with $\phi_s = (N-1)^{-1}(NRes_{(N;1)s} - 1)$. As $\mu_N(\mathcal{Y})$ forms a basis for $\langle \mu_N(\mathcal{C}) \rangle$ then for any $Z \in \langle \mathcal{C} \rangle$ we have

$$\mu_N(Z) = \sum_{s=1}^{v_0} \text{Cov}\{\mu_N(Z), \mu_N(Y_s)\}r_{(N;0)s}\mu_N(Y_s) = \sum_{s=1}^{v_0} \text{Cov}\{\mu(Z), \mu(Y_s)\}r_{0s}\mu_N(Y_s)$$

so that it is straightforward to obtain the adjusted mean and variance of $\mu_N(Z)$. The collection $\mu_N(\mathcal{Y})$ form an orthogonal grid over $\langle \mu_N(\mathcal{C}) \rangle$ for which we expect to learn most, in terms of variance reduction, about those quantities with large correlations with the early $\mu_N(Y_s)$. The impact of the sample size upon the quantitative features of the update are thus easy to assess. For example, if we are interested in choosing a sample size to achieve a specified variance reduction over elements in $\langle \mu_N(\mathcal{C}) \rangle$ then we have the following corollary to aid us in this process.

Corollary 1 *For any $\mu_N(Y_s) \in \mu_N(\mathcal{Y})$, the sample size n required to achieve a proportional variance reduction of $0 < \kappa < 1$ for $\mu_N(Y_s)$, that is so that $\text{Var}_{\mathcal{C}(n)}\{\mu_N(Y_s)\} \leq (1 - \kappa)\text{Var}\{\mu_N(Y_s)\}$, is $n \geq N\kappa r_{(N;0)s} / \{(N - 1)(1 - \kappa)r_{(N)s} + \kappa r_{(N;0)s}\}$. If $\phi_{\bar{s}}$ is the minimal eigenvalue then to achieve a proportionate variance reduction of κ for every element of $\langle \mu_N(\mathcal{C}) \rangle$ we require a sample size, rounded up, of $N\kappa r_{(N;0)\bar{s}} / \{(N - 1)(1 - \kappa)r_{(N)\bar{s}} + \kappa r_{(N;0)\bar{s}}\}$.*

Note that, in addition to not depending upon n , the orthogonal grid formed by $\mu_N(\mathcal{Y})$ shares the same coordinate representation for each N and also to the orthogonal grid $\mu(\mathcal{Y})$ forms over $\langle \mu(\mathcal{C}) \rangle$. We need only solve the generalised eigenvalue problem given in Definition 2 to obtain the solution for any choice of n and N . Irrespective of sample size or population size the qualitative features of the update are the same whilst it is straightforward to assess the impact of the population size upon the quantitative features of the update. We illustrate this by forming a direct comparison between the results of Theorems 1 and 2. By first obtaining $\frac{r_{0s}}{r_{ns}}$ from (12) and $\frac{r_{(N;0)s}}{r_{(N;n)s}}$ from (16) we have

$$\frac{r_{(N;0)s}}{r_{(N;n)s}} = \left(1 - \frac{n}{N}\right) \frac{r_{0s}}{r_{ns}}. \quad (17)$$

Using (17) we may express (14) as

$$\mu_{(N;n)s} = \left(1 - \frac{n}{N}\right) \mu_{ns} + \frac{n}{N} \bar{Y}_s. \quad (18)$$

Equation (17) shows that, for each $s = 1, \dots, v_0$, the ratio of adjusted to prior variance of $\mu_N(Y_s)$ is equal to the ratio of adjusted to prior variance of $\mu(Y_s)$ multiplied by the finite population correction term $1 - \frac{n}{N}$. For the corresponding expectations, (18) reveals that the expectation of each $\mu_N(Y_s)$ given $\mathcal{C}(n)$ is a weighted average of the expectation of $\mu(Y_s)$ given $\mathcal{C}(n)$ and the observed mean \bar{Y}_s , the weights being dependent upon the ratio of the total population observed in the sample.

3.4 Extendible second-order exchangeable populations

As we discussed in §3.1, a finite sequence could form part of a larger (possibly infinite) sequence of second-order exchangeable collections or it may not be embedded in any longer sequence. The results of Theorem 2, derived entirely from beliefs over observable random quantities, hold irrespective of whether the finite sequence may be embedded in a larger one. It is, however, natural to consider the circumstances when a second-order exchangeable sequence may be embedded, or extended, into a longer sequence of similarly defined random quantities.

Definition 3 *Suppose that the collection of measurements \mathcal{C} is second-order exchangeable over the coherently specified $\mathcal{C}_N^* = \cup_{i=1}^N \mathcal{C}_i$. The population \mathcal{C}_N^* is q -extendible if \mathcal{C} is also second-order exchangeable over $\mathcal{C}_{N+q}^* = \cup_{i=1}^{N+q} \mathcal{C}_i$, where \mathcal{C}_{N+q}^* is coherently specified.*

For the original sequence and the extended sequence, our judgements for each individual and between each pair of individuals are the same. Infinite exchangeability may be viewed as q -extendibility for all $q > 0$. If $\text{Cov}(Y_i, Y_j) \geq 0$ for all $Y \in \langle \mathcal{C} \rangle$ then we have q -extendibility for all $q > 0$. Otherwise, we have q -extendibility for all $q \leq 1 - N - \rho_{\min}^{-1}$ where $\rho_{\min} = \{\min_{Y \in \langle \mathcal{C} \rangle} \text{Corr}(Y_i, Y_j) : \text{Cov}(Y_i, Y_j) < 0\}$. Recall that, for \mathcal{C} to be second-order exchangeable over \mathcal{C}^* , we are required only to specify the relationship for \mathcal{C}_2^* with all other cases following by symmetry. Thus, we can regard any second-order exchangeable sequence as having been extended from a second-order sequence of length two and the effect of a sample of size n deduced from a sample of size one. Noting that, from (15), we can deduce that $r_{(N;n)s} = g(N, n)nr_{(N;1)s} - \frac{N(n-1)}{N-n}r_{(N;0)s}$ and $\frac{r_{(N;1)s}}{r_{(N;0)s}} = \frac{N}{2(N-1)} \frac{r_{(2;1)s}}{r_{(1;0)s}}$. Using these identities we can obtain the following corollary to Theorem 2 which shows how to extend the results from a population of size two and a sample of size one to any finite population and sample size.

Corollary 2 *The orthogonal collections $\mu_2(\mathcal{Y})$ and $\mu_N(\mathcal{Y})$ share the same coordinate representation with*

$$\begin{aligned} \frac{r_{(N;n)s}}{r_{(N;0)s}} &= \left(1 - \frac{n}{N}\right)^{-1} \left\{ \frac{n}{2} \frac{r_{(2;1)s}}{r_{(2;0)s}} - (n-1) \right\}, \\ \mu_{(N;n)s} &= \left(1 - \frac{n}{N}\right) \left\{ n - 2(n-1) \frac{r_{(2;0)s}}{r_{(2;1)s}} \right\}^{-1} \{ 2\mu_{(2;1)s} - Y_{s1} - \bar{Y}_{s-} \} + \\ &\quad \frac{1}{N} \{ Y_{s1} + (N-1)\bar{Y}_{s-} \}, \end{aligned} \quad (19)$$

where $\bar{Y}_{s-} = \frac{1}{n-1} \sum_{i=2}^n Y_{si}$ for $n > 1$ and zero otherwise.

The qualitative information provided by the adjustment of $\langle \mu_N(\mathcal{C}) \rangle$ by $\mathcal{C}(n)$ remains the same for all possible sequence lengths N and all possible sample sizes n and the quantitative information is easy to compare across these via equations (19) and (20). Equations (17) and (18) show how the use of a fpc can be used to compare the finite and infinite cases. Analogous results may be obtained if we wish to compare two finite population sizes, N_1 and N_2 say, which illustrate the fundamental roles played by the two sampling fractions, $\frac{n}{N_1}$ and $\frac{n}{N_2}$. From (19) we have

$$\left(1 - \frac{n}{N_1}\right) \frac{r_{(N_1;n)s}}{r_{(N_1;0)s}} = \left(1 - \frac{n}{N_2}\right) \frac{r_{(N_2;n)s}}{r_{(N_2;0)s}},$$

whilst (20) gives

$$\mu_{(N_1;n)s} = \left(1 - \frac{n}{N_1}\right) \left(1 - \frac{n}{N_2}\right)^{-1} \mu_{(N_2;n)s} + \left(\frac{n}{N_1} - \frac{n}{N_2}\right) \left(1 - \frac{n}{N_2}\right)^{-1} \bar{Y}_s.$$

This section has illustrated how the familiar univariate fpc naturally generalises to individual quantities in the multivariate population and these quantities, for any population size and sample size, are derived from the underlying canonical variable directions obtained in Definition 2. The qualitative information may be derived from the corresponding underlying canonical variable resolutions. Not only is this theoretically important but there is a huge computational advantage: for any choice of n and N we are only required to solve the single $v_0 \times v_0$ generalised eigenvalue problem given in Definition 2.

4 Example: summarising exchangeable observations

4.1 General framework

To aide the illustration of the theory we shall make a further assumption which enables the explicit derivation of the underlying canonical variable directions and resolutions.

Assumption 2 *For each individual we judge that the v_0 measurements $\mathcal{C}_i = \{X_{1i}, \dots, X_{v_0i}\}$ under the model given by Assumption 1 are second-order exchangeable and co-exchangeable across individuals. In this case, the model for all $i \neq j$ reduces to*

$$E(\mathcal{C}_i) = \mu_0 \mathbf{1}_{v_0}; \text{ Var}(\mathcal{C}_i) = (d_1 - d_2)I_{v_0} + d_2 J_{v_0}; \text{ Cov}(\mathcal{C}_i, \mathcal{C}_j) = (c_1 - c_2)I_{v_0} + c_2 J_{v_0}$$

where μ_0, d_1, d_2, c_1 , and c_2 are constants, $\mathbf{1}_{v_0}$ is the $v_0 \times 1$ vector of 1s, I_{v_0} the $v_0 \times v_0$ identity matrix and $J_{v_0} = \mathbf{1}_{v_0} \mathbf{1}_{v_0}^T$.

Under this model we have $R^{-1} = \{(d_1 - d_2) - (c_1 - c_2)\}I_{v_0} + (d_2 - c_2)J_{v_0}$ and $R_0^{-1} = (c_1 - c_2)I_{v_0} + c_2 J_{v_0}$. Hence, the matrices R and R_0 have a particularly simple form ($aI_{v_0} + bJ_{v_0}$ for constants a and b) so that, see Definition 2, the eigenstructure of the problem $RH = (R + R_0)H\Phi$ is analytically straightforward to obtain. We explore the use of the fpc when the population is of size N . The eigenvalues of (10) are given by

$$\phi_1 = \frac{c_1 + (v_0 - 1)c_2}{d_1 + (v_0 - 1)d_2}; \phi_2 = \dots = \phi_{v_0} = \frac{c_1 - c_2}{d_1 - d_2}. \quad (21)$$

Notice that $\phi_1 \geq \phi_2 \Leftrightarrow c_2 d_1 \geq d_2 c_1$ so that the ordering of the eigenvalues will depend upon our specific prior choice of d_1, d_2, c_1 and c_2 . The eigenvector corresponding to ϕ_1 is proportional to $\mathbf{1}_{v_0}$ and the eigenvectors corresponding to ϕ_2 are any $v_0 - 1$ orthogonal vectors which are orthogonal to $\mathbf{1}_{v_0}$ and whose coefficients sum to zero. Thus, the first eigenvector is an average and the remainder are $v_0 - 1$ linear contrasts. For simplicity of exposition, we will take $\Phi = \text{diag}(\phi_1, \phi_2, \dots, \phi_{v_0})$ to be our matrix of eigenvalues with the corresponding matrix of eigenvectors, normed as in Definition 2, given by $H = H_{v_0} \tilde{\Phi}$ where

$$\tilde{\Phi} = \text{diag}(\sqrt{\{d_1 + (v_0 - 1)d_2\}(1 - \phi_1)}, \sqrt{(d_1 - d_2)(1 - \phi_2)}, \dots, \sqrt{(d_1 - d_2)(1 - \phi_{v_0})})$$

and H_{v_0} is the $v_0 \times v_0$ transpose of the Helmert matrix of order v_0 . The first column of H_{v_0} is $\frac{1}{\sqrt{v_0}} \mathbf{1}_{v_0}$ and, for $v > 1$, the v th column is $\frac{1}{\sqrt{v(v-1)}} (-1_{v-1}^T \ v - 1 \ 0 \dots 0)^T$. The underlying canonical variable directions are the columns of the matrix $W = RH = H_{v_0} \tilde{\Phi}^{-1}$. We then form the collection $\mu_N(\mathcal{Y}) = \{\mu_N(Y_1), \dots, \mu_N(Y_{v_0})\}$ where

$$\mu_N(Y_1) = \alpha_1 \sum_{v=1}^{v_0} \mu_N(X_v); \quad (22)$$

$$\mu_N(Y_s) = \alpha_s \left\{ (s-1)\mu_N(X_s) - \sum_{v=1}^{s-1} \mu_N(X_v) \right\}, \quad s = 2, \dots, v_0, \quad (23)$$

and $\alpha_1 = (\sqrt{v_0} \tilde{\Phi}_{11})^{-1}$, $\alpha_s = (\sqrt{s(s-1)} \tilde{\Phi}_{22})^{-1}$. From Theorem 2, the familiar univariate fpc attaches itself to each of the quantities given by (22) and (23). We shall explicitly illustrate the construction for the precisions given by (15); the illustration for the means given by (14) is obtained in a similar fashion. For each $s = 1, \dots, v_0$ we find the prior precision, $r_{(N;0)s}$, of $\mu_N(Y_s)$ and the posterior adjusted precision $r_{(N)s}$ of Y_{si} given $\mu_N(\mathcal{C})$

where $Y_{1i} = \alpha_1 \sum_{v=1}^{v_0} X_{vi}$ and $Y_{si} = \alpha_s \{(s-1)X_{vi} - \sum_{v=1}^{s-1} X_{vi}\}$ for $s = 2, \dots, v_0$. We have

$$r_{(N;0)1} = \frac{N\{(d_1 - c_1) + (v_0 - 1)(d_2 - c_2)\}}{\{d_1 + (v_0 - 1)d_2\} + (N-1)\{c_1 + (v_0 - 1)c_2\}}; \quad (24)$$

$$r_{(N;0)s} = \frac{N\{(d_1 - d_2) - (c_1 - c_2)\}}{(d_1 - d_2) + (N-1)(c_1 - c_2)}, \quad s = 2, \dots, v_0, \quad (25)$$

and $r_{(N)s} = \frac{N}{N-1}$ for $s = 1, \dots, v_0$. For each s , the posterior precision, $r_{(N;n)s}$ of $\mu_N(Y_s)$ is the sum of the prior precision plus the precision of the sample mean, with the latter corrected by the fpc $g(N, n)$, that is $r_{(N;n)s} = r_{(N;0)s} + g(N, n)nr_{(N)s}$. Explicitly computing each $r_{(N;n)s}$ we have

$$r_{(N;n)1} = \frac{N^2}{N-n} \times \frac{\{d_1 + (v_0 - 1)d_2\} + (n-1)\{c_1 + (v_0 - 1)c_2\}}{\{d_1 + (v_0 - 1)d_2\} + (N-1)\{c_1 + (v_0 - 1)c_2\}}; \quad (26)$$

$$r_{(N;n)s} = \frac{N^2}{N-n} \times \frac{(d_1 - d_2) + (n-1)(c_1 - c_2)}{(d_1 - d_2) + (N-1)(c_1 - c_2)}, \quad s = 2, \dots, v_0. \quad (27)$$

Observe that by dividing (26) by (24) and (27) by (25) and then multiplying each by $(1 - \frac{n}{N})$ we have that, for each $s = 1, \dots, v_0$, $(1 - \frac{n}{N})\frac{r_{(N;n)s}}{r_{(N;0)s}}$ does not depend upon N as we expected from (17). Notice that, from Theorem 2, the $\mu_N(Y_s)$ are a priori and a posteriori uncorrelated so that we can use them to learn about any quantity in $\langle \mu_N(\mathcal{C}) \rangle$. As an illustration, from (22) and (23) we may obtain that

$$\begin{aligned} \mu_N(X_1) &= \frac{\alpha_1}{v_0} \mu_N(Y_1) - \sum_{v=2}^{v_0} \frac{\alpha_v}{v(v-1)} \mu_N(Y_v); \\ \mu_N(X_s) &= \frac{\alpha_1}{v_0} \mu_N(Y_1) + \frac{\alpha_s}{s} \mu_N(Y_s) - \sum_{v=s+1}^{v_0} \frac{\alpha_v}{v(v-1)} \mu_N(Y_v), \quad s = 2, \dots, v_0, \end{aligned}$$

so that

$$\begin{aligned} \text{Var}_{\mathcal{C}(n)}\{\mu_N(X_1)\} &= \frac{\alpha_1^2}{v_0^2} r_{(N;n)1}^{-1} + \sum_{v=2}^{v_0} \frac{\alpha_v^2}{v^2(v-1)^2} r_{(N;n)v}^{-1}; \\ \text{Var}_{\mathcal{C}(n)}\{\mu_N(X_s)\} &= \frac{\alpha_1^2}{v_0^2} r_{(N;n)1}^{-1} + \frac{\alpha_s^2}{s^2} r_{(N;n)s}^{-1} + \sum_{v=s+1}^{v_0} \frac{\alpha_v^2}{v^2(v-1)^2} r_{(N;n)v}^{-1}, \end{aligned}$$

for $s = 2, \dots, v_0$. We now illustrate, via an example concerning examination data, how we can use these results for setting appropriate sample sizes.

4.2 Examination data

In a similar vein to §5 of Goldstein (1988), we consider an examination sat by 854 candidates. Each candidate was required to answer six compulsory questions, each marked out of 10. Any question not attempted received a mark of zero. The chief examiner has a number of questions of interest such as whether the questions were of roughly similar difficulty and if the standard was similar to previous years and will take a sample of the 854 scripts to help answer these. Let X_{vi} denote the mark of candidate i on question v . $\mu_{854}(X_v) = \frac{1}{854} \sum_{i=1}^{854} X_{vi}$ is thus the average mark on question v . The chief examiner judges that the candidates are second-order exchangeable and also that the questions

are second-order exchangeable so that he follows the model given in Assumption 2 and thus the framework of §4.1. He judges that the expected score on each question is 6.5, that is $\mu_0 = 6.5$. He wishes to explore a range of values for his prior variance statements and their impact upon his choice of sample size. He chooses $c_1 = 1$ and considers $0 < c_2 < 0.95$ so that, if the population were infinite, c_1 would reflect the variance of the underlying average mark for each question and c_2 the covariance between the underlying average marks for different questions. He wishes to explore $2 < d_1 < 4$, so that $1 < d_1 - c_1 < 3$: larger values of d_1 representing greater candidate variation. Finally, he sets $d_2 = 2c_2$ so that $d_2 - c_2 = c_2$.

The chief examiner solves the generalised eigenvalue problem given by Definition 2. The canonical variable resolutions have the form given by (21). Notice that for $d_1 > 2$ we have $\phi_1 > \phi_2$ with equality when $d_1 = 2$. The collection $\mu_{854}(\mathcal{Y})$ may be expressed by (22) and (23) (with $v_0 = 6$). Thus, $\mu_{854}(Y_1)$ is proportional to the overall mark on the examination and this enables the chief examiner to learn about the general standard of the exam, a high/low value suggesting an easy/hard examination. The collection $\mu_{854}(Y_2), \dots, \mu_{854}(Y_6)$ may be viewed as summarising all of the differences between the difficulty of the questions.

The chief examiner is interested in selecting the sample size n to achieve a proportional variance reduction of κ for the overall mark in the examination. This is equivalent to choosing the n to achieve this task for $\mu_{854}(Y_1)$ and the n may be found using Corollary 1. Goldstein and Wooff (1997) provide an overview of how sample size selection can be performed for experiments analysed using a Bayes linear approach assuming an infinite population. In Figure 1(a)-(c), the values of n are plotted for three values of κ . We observe that as d_1 increases so does n which is to be expected due to the increasing candidate variability. As c_2 decreases, n also increases which is again expected as this controls both the covariance between questions and candidates. The most extreme case of $d_1 = 4$ and $c_2 = 0$ would require a sample size of 27, which is about 3% of the population, to achieve a variance reduction of 0.9. Observe that we need to approximately double the sample size to increase the variance reduction from 0.90 to 0.95. A sample size of 220 is required for a reduction of 0.99 in the most extreme case which is approximately 26% of the population.

If instead, the chief examiner is interested in selecting the sample size n to achieve a proportional variance reduction of κ for learning about the differences between the questions then this is equivalent to choosing the n to achieve this task for $\mu_{854}(Y_s)$, $s = 2, \dots, 6$. Once again, Corollary 1 may be used to find the n and Figure 1(d)-(f) shows the values of n for three values of κ . Notice that, for any value of κ , if $d_1 = 2$ the sample size required is the same as that for $\mu_{854}(Y_1)$ and that this is where $\phi_1 = \phi_2$. For all other plotted values of d_1 and c_2 the sample size required is larger than that for $\mu_{854}(Y_1)$ which is to be expected as here $\phi_2 < \phi_1$. In some cases, it may be appreciably higher for example when d_1 is large and c_2 is large which corresponds to large candidate variation and high correlation between questions making it difficult to learn about the differences.

Finally, it is important to note here that as there are only two canonical variable resolutions, the plots give upper and lower bounds on the sample size n required to achieve a proportional variance reduction of κ for any linear combination of the $\mu_{854}(X_v)$ of interest to the chief examiner.

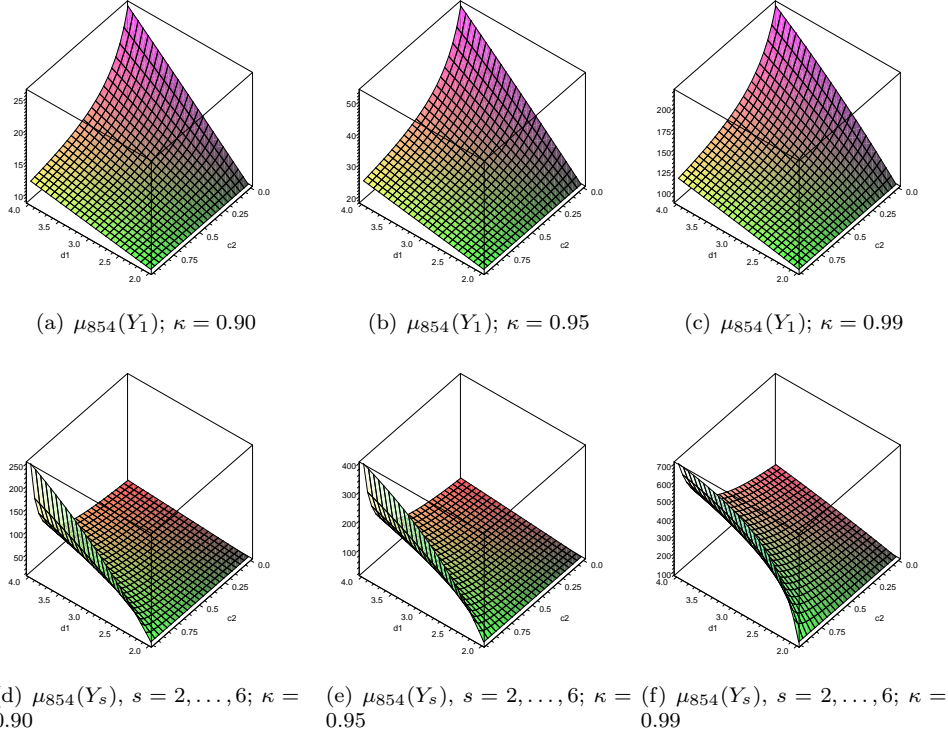


Figure 1: Sample size n required to achieve a proportional variance reduction of κ for $\mu_{854}(Y_1)$ and each $\mu_{854}(Y_s)$, $s = 2, \dots, 6$ for the examination problem with $c_1 = 1$, $0 < c_2 < 0.95$, $2 < d_1 < 4$ and $d_2 = 2c_2$.

5 Cluster sampling

We now show how the results of §3 can be utilised when we extend the model of second-order exchangeability, as given in Assumption 1, to encompass two-stage cluster sampling where each individual additionally belongs to a cluster. We intend to make the same series of measurements $\mathcal{C} = \{X_1, \dots, X_{v_0}\}$ on each individual and let $\mathcal{C}_{gi} = \{X_{g1i}, \dots, X_{gv_0i}\}$ denote the measurements for the i th individual in the g th cluster and we suppose that there are a total of M clusters each of which contain N individuals.

Assumption 3 *We judge that individuals in each cluster are second-order exchangeable and that they are co-exchangeable (Goldstein, 1986) across clusters. For all $g \neq h$, $i \neq j$, k our second-order specifications thus take the form*

$$E(\mathcal{C}_{gi}) = \mu_g; \text{Var}(\mathcal{C}_{gi}) = D_g; \text{Cov}(\mathcal{C}_{gi}, \mathcal{C}_{gj}) = C_{gg}; \text{Cov}(\mathcal{C}_{gi}, \mathcal{C}_{hk}) = C_{gh}.$$

Using the second-order representation theorem of Goldstein (1986) we may write

$$\mathcal{C}_{gi} = \mu_N(\mathcal{C}_g) + R_{N,i}(\mathcal{C}_g) \quad (28)$$

where $\mu_N(\mathcal{C}_g) = \frac{1}{N} \sum_{i=1}^N \mathcal{C}_{gi}$ is the g th cluster population mean vector and $R_{N,i}(\mathcal{C}_g) = \mathcal{C}_{gi} - \mu_N(\mathcal{C}_g)$ the residual vector for the i th individual in the g th cluster. For all g, h , i , $R_{N,i}(\mathcal{C}_g)$ is uncorrelated with $\mu_N(\mathcal{C}_h)$. We collect the cluster population mean vectors together as $\mu_N(\mathcal{C}) = \{\mu_N(\mathcal{C}_1), \dots, \mu_N(\mathcal{C}_M)\}$.

Assumption 4 We judge that the cluster population means are second-order exchangeable. For all $g \neq h$ our second-order specifications thus take the form

$$\begin{aligned} E(\mu_N(\mathcal{C}_g)) &= \mu_g = \mu; \\ \text{Var}(\mu_N(\mathcal{C}_g)) &= C_{gg} + \frac{1}{N}(D_g - C_{gg}) = A; \\ \text{Cov}(\mu_N(\mathcal{C}_g), \mu_N(\mathcal{C}_h)) &= C_{gh} = B \end{aligned}$$

where the vector μ and matrices A and B do not depend upon g .

From the second-order representation theorem of Goldstein (1986) we may write

$$\mu_N(\mathcal{C}_g) = \mu_{(M;N)}(\mathcal{C}) + R_{(M;N),g}(\mathcal{C}) \quad (29)$$

where $\mu_{(M;N)}(\mathcal{C}) = \frac{1}{M} \sum_{g=1}^M \mu_N(\mathcal{C}_g)$ is the overall population mean and $R_{(M;N),g}(\mathcal{C}) = \mu_N(\mathcal{C}_g) - \mu_{(M;N)}(\mathcal{C})$. For each g , $R_{(M;N),g}(\mathcal{C})$ is uncorrelated with $\mu_{(M;N)}(\mathcal{C})$. Combining the two representations given by (28) and (29) we may write

$$\mathcal{C}_{gi} = \mu_{(M;N)}(\mathcal{C}) + R_{(M;N),g}(\mathcal{C}) + R_{N,i}(\mathcal{C}_g) \quad (30)$$

where, for each i, g , the three components in the decomposition of \mathcal{C}_{gi} are mutually uncorrelated. However, due to the finite nature of the populations, the $R_{(M;N),g}(\mathcal{C})$ are not uncorrelated across clusters but rather, as Goldstein (1986; p974-975) terms, uncorrelated to order $\frac{1}{M}$ whilst the $R_{N,i}(\mathcal{C}_g)$ are uncorrelated across clusters and uncorrelated to order $\frac{1}{N}$ within clusters. In a univariate setting, models derived from decompositions analogous to (30), though with different stochastic structure, have been studied by Stanek and Singer (2004) and Martino et al (2008). One such example is that of Scott and Smith (1969), see also Little and Zheng (2007), though this explicitly has observations in differing clusters being uncorrelated whereas our use of co-exchangeability allows a correlation between observations in differing clusters.

Suppose that we sample m of the M clusters and in each of the m sampled clusters, sample n individuals. For notational simplicity we use the labelling convention that we sample the first m clusters and, in each sampled cluster, the first n individuals. Let $\mathcal{C}_g(n) = \{\mathcal{C}_{g1}, \dots, \mathcal{C}_{gn}\}$ denote the measurements of the individuals sampled in the g th cluster and $\mathcal{C}(m; n) = \{\mathcal{C}_1(n), \dots, \mathcal{C}_m(n)\}$ the total collection of observations. The observed sample mean in the g th cluster is $\bar{\mathcal{C}}_g = \frac{1}{n} \sum_{i=1}^n \mathcal{C}_{gi}$ so that the total collection of sample means is $\bar{\mathcal{C}}(m; n) = \{\bar{\mathcal{C}}_1, \dots, \bar{\mathcal{C}}_m\}$. Theorems 1 and 2 show that only the observed sample mean was sufficient to compute the posterior quantities. In this extended model, we can restrict attention to various sample means by exploiting the concept of Bayes linear sufficiency (Goldstein and O'Hagan, 1996). We have the following lemma; the proof is in the appendix.

Lemma 1 1. The collection of sample means $\bar{\mathcal{C}}(m; n)$ is Bayes linear sufficient for the observations $\mathcal{C}(m; n)$ for adjusting the overall mean $\mu_{(M;N)}(\mathcal{C})$, so that

$$\begin{aligned} E_{\mathcal{C}(m; n)}(\mu_{(M;N)}(\mathcal{C})) &= E_{\bar{\mathcal{C}}(m; n)}(\mu_{(M;N)}(\mathcal{C})); \\ \text{Var}_{\mathcal{C}(m; n)}(\mu_{(M;N)}(\mathcal{C})) &= \text{Var}_{\bar{\mathcal{C}}(m; n)}(\mu_{(M;N)}(\mathcal{C})). \end{aligned}$$

2. The population mean across the sampled clusters, $\mu_{(m;N)}(\mathcal{C}) = \frac{1}{m} \sum_{g=1}^m \mu_N(\mathcal{C}_g)$, is Bayes linear sufficient for $\bar{\mathcal{C}}(m; n)$ for adjusting $\mu_{(M;N)}(\mathcal{C})$, so that

$$\begin{aligned} E_{\bar{\mathcal{C}}(m; n)}(\mu_{(M;N)}(\mathcal{C})) &= E_{\bar{\mathcal{C}}(m; n)}\{E_{\mu_{(m;N)}(\mathcal{C})}(\mu_{(M;N)}(\mathcal{C}))\}; \\ \text{Var}_{\bar{\mathcal{C}}(m; n)}(\mu_{(M;N)}(\mathcal{C})) &= \text{Var}_{\mu_{(m;N)}(\mathcal{C})}(\mu_{(M;N)}(\mathcal{C})) \\ &\quad + \text{Var}_{\bar{\mathcal{C}}(m; n)}\{E_{\mu_{(m;N)}(\mathcal{C})}(\mu_{(M;N)}(\mathcal{C}))\}. \end{aligned}$$

Lemma 1 shows that the adjustment can be performed in two stages. In the first, the population mean $\mu_{(M;N)}(\mathcal{C})$ is adjusted following observation of $\mu_{(m;N)}(\mathcal{C})$, the mean across the sampled clusters, and so the adjustment is related to the proportion of clusters sampled. In the second stage $\mu_{(m;N)}(\mathcal{C})$ is adjusted by the sample means $\bar{\mathcal{C}}(m; n)$ and thus relates to the proportion of individuals sampled in the sampled clusters. Notice that this separation mirrors that in the (univariate) classical setting, see for example Chapter 10 of Cochran (1977). We now consider separately each of these stages, showing that each can be viewed as a finite population sampling problem, before combining them together to find the full adjustment.

5.1 Sampling m from M : adjustment of $\langle \mu_{(M;N)}(\mathcal{C}) \rangle$ given $\mu_{(m;N)}(\mathcal{C})$

We now consider the adjustment of quantities contained in $\langle \mu_{(M;N)}(\mathcal{C}) \rangle$ given $\mu_{(m;N)}(\mathcal{C})$ under the model formed by Assumptions 3 and 4. From the model, $\mu_N(\mathcal{C}(M)) = \{\mu_N(\mathcal{C}_1), \dots, \mu_N(\mathcal{C}_M)\}$ is a second-order exchangeable population of size M and, from the representation given by (29), $\mu_{(M;N)}(\mathcal{C})$ is the underlying population mean. If we observe m of the M population members, $\mu_N(\mathcal{C}(m)) = \{\mu_N(\mathcal{C}_1), \dots, \mu_N(\mathcal{C}_m)\}$, then the sample mean $\mu_{(m;N)}(\mathcal{C}) = \frac{1}{m} \sum_{g=1}^m \mu_N(\mathcal{C}_g)$ is Bayes linear sufficient for $\mu_N(\mathcal{C}(m))$ for adjusting $\langle \mu_{(M;N)}(\mathcal{C}) \rangle$.

Definition 4 *The underlying canonical cluster directions are defined as the columns of the matrix $U = [U_1 \dots U_{v_0}]$ where $U_s = SF_s$ and F_s is the s th column of the matrix F solving the generalised eigenvalue problem*

$$SF = (S + S_0)F\Psi$$

where $\Psi = \text{diag}(\psi_1, \dots, \psi_{v_0})$ is the matrix of eigenvalues and, for $g \neq h$, $S^{-1} = \text{Var}(\mu_N(\mathcal{C}_g)) - \text{Cov}(\mu_N(\mathcal{C}_g), \mu_N(\mathcal{C}_h))$ and $S_0^{-1} = \text{Cov}(\mu_N(\mathcal{C}_g), \mu_N(\mathcal{C}_h))$. F is normed so that $F^T S F = I$ and $F^T (S + S_0) F \Psi = I$. The ordered eigenvalues $1 > \psi_1 \geq \dots \geq \psi_{v_0} > 0$ are termed the underlying canonical cluster resolutions.

Let U_{vs} denote the v th component of the s th canonical cluster direction and, for each $s = 1, \dots, v_0$, define $W_s \in \langle \mathcal{C} \rangle$ to be $W_s = \sum_{v=1}^{v_0} U_{vs} X_v$. The corresponding quantities in $\langle \mu_{(M;N)}(\mathcal{C}) \rangle$, $\langle \mu_{(m;N)}(\mathcal{C}) \rangle$ and each $\langle \mu_N(\mathcal{C}_g) \rangle$ which share the same coordinate representation as W_s are defined to be $\mu_{(M;N)}(W_s)$, $\mu_{(m;N)}(W_s)$ and $\mu_N(W_{gs})$ respectively. An application of Theorem 2 gives the following corollary.

Corollary 3 *For a sample of size m drawn from a population of size M , the collection $\mu_{(M;N)}(\mathcal{W}) = \{\mu_{(M;N)}(W_1), \dots, \mu_{(M;N)}(W_{v_0})\}$ forms a basis for $\langle \mu_{(M;N)}(\mathcal{C}) \rangle$. The $\mu_{(M;N)}(W_s)$ are a priori uncorrelated, and, for all samples of any size, a posteriori uncorrelated. The posterior adjusted expectation, $\mu_{[M;m]s} = E_{\mu_{(m;N)}(\mathcal{C})}(\mu_{(M;N)}(W_s))$, and posterior adjusted precision, $r_{[M;m]s} = \text{Var}_{\mu_{(m;N)}(\mathcal{C})}^{-1}(\mu_{(M;N)}(W_s))$, for $\mu_{(M;N)}(W_s)$ are given by*

$$\mu_{[M;m]s} = \frac{r_{[M;0]s}\mu_{[M;0]s} + g(M, m)mr_{[M]s}\mu_{(m;N)}(W_s)}{r_{[M;0]s} + g(M, m)mr_{[M]s}}; \quad (31)$$

$$r_{[M;m]s} = r_{[M;0]s} + g(M, m)mr_{[M]s}, \quad (32)$$

where $\mu_{[M;0]s}$, $r_{[M;0]s}$ are the prior expectation and precision for $\mu_{(M;N)}(W_s)$, $r_{[M]s}$ is the adjusted precision for $\mu_N(W_{gs})$ given $\mu_{(M;N)}(\mathcal{C})$, and $g(M, m)$ the fpc.

The canonical cluster directions and resolutions, for any choice of m and M given N , completely summarise the adjustment of $\langle \mu_{(M;N)}(\mathcal{C}) \rangle$ given $\mu_{(m;N)}(\mathcal{C})$ and, from Lemma 1, give the first stage of the adjustment of $\langle \mu_{(M;N)}(\mathcal{C}) \rangle$ given $\mathcal{C}(m; n)$.

5.2 Sampling n from N : adjustment of $\langle \mu_{(m;N)}(\mathcal{C}) \rangle$ given $\mu_{(m;n)}(\mathcal{C})$

The second stage of the adjustment of $\langle \mu_{(M;N)}(\mathcal{C}) \rangle$ given $\mathcal{C}(m; n)$ requires the adjustment of $\langle \mu_{(m;N)}(\mathcal{C}) \rangle$ given $\bar{\mathcal{C}}(m; n)$. We can exploit the symmetries in our beliefs, as given by Assumptions 3 and 4, to invert the variance matrix of $\bar{\mathcal{C}}(m; n)$ and, utilising the resolution transform (see §3.9 of Goldstein and Wooff, 2007), reduce this latter problem to one of solving the generalised eigenvalue problem

$$\left\{ B + \frac{1}{m}(A - B) \right\} V_{(n)} = (\Omega^{-1} + B) V_{(n)} \Xi_{(n)}$$

where $\Omega = \sum_{g=1}^m \{ (C_{gg} - B) + \frac{1}{n}(D_g - C_{gg}) \}^{-1}$. Typically both $V_{(n)}$, the matrix of eigenvectors, and $\Xi_{(n)}$, the diagonal matrix of eigenvalues, will depend upon the within cluster population size N and sample size n in a less than tractable way. An additional assumption that the $\bar{\mathcal{C}}(m; n)$ are second-order exchangeable reduces the problem to one of sampling n individuals from a finitely second-order exchangeable population of size N .

Lemma 2 *If the collection of sample means $\bar{\mathcal{C}}(m; n)$ is second-order exchangeable then the overall observed mean $\mu_{(m;n)}(\mathcal{C}) = \frac{1}{m} \sum_{g=1}^m \bar{\mathcal{C}}_g$ is Bayes linear sufficient for $\bar{\mathcal{C}}(m; n)$ for adjusting $\mu_{(m;N)}(\mathcal{C})$, so that*

$$\begin{aligned} E_{\bar{\mathcal{C}}(m;n)}(\mu_{(m;N)}(\mathcal{C})) &= E_{\mu_{(m;n)}(\mathcal{C})}(\mu_{(m;N)}(\mathcal{C})); \\ \text{Var}_{\bar{\mathcal{C}}(m;n)}(\mu_{(m;N)}(\mathcal{C})) &= \text{Var}_{\mu_{(m;n)}(\mathcal{C})}(\mu_{(m;N)}(\mathcal{C})). \end{aligned}$$

Proof - Using the second-order representation theorem for each g we may write $\bar{\mathcal{C}}_g = \mu_{(m;n)}(\mathcal{C}) + R_{(m;n),g}(\mathcal{C})$ where $R_{(m;n),g}(\mathcal{C}) = \bar{\mathcal{C}}_g - \mu_{(m;n)}(\mathcal{C})$ is orthogonal to $\mu_{(m;n)}(\mathcal{C})$ so that $\text{Cov}(\mu_{(m;n)}(\mathcal{C}), \bar{\mathcal{C}}_g) = \text{Var}(\mu_{(m;n)}(\mathcal{C}))$. As $\text{Cov}(\mu_{(m;N)}(\mathcal{C}), \bar{\mathcal{C}}_h) = \text{Var}(\mu_{(m;N)}(\mathcal{C}))$ it immediately follows that $\text{Cov}(\mu_{(m;N)}(\mathcal{C}), \text{Cov}(\mu_{(m;n)}(\mathcal{C})) = \text{Var}(\mu_{(m;N)}(\mathcal{C}))$. Hence, $\text{Cov}(\mu_{(m;N)}(\mathcal{C}), \mu_{(m;n)}(\mathcal{C})) \text{Var}^{-1}(\mu_{(m;n)}(\mathcal{C})) \text{Cov}(\mu_{(m;n)}(\mathcal{C}), \bar{\mathcal{C}}(m; n)) = \text{Cov}(\mu_{(m;N)}(\mathcal{C}), \bar{\mathcal{C}}(m; n))$ and the result follows from Theorems 5.20 and 5.23 of Goldstein and Wooff (2007). \square

The effect of the additional judgement that $\bar{\mathcal{C}}(m; n)$ is exchangeable is that the the matrices D_g and C_{gg} do not depend upon g and so, for all $g \neq h$, $i \neq j$, k , our second-order specifications take the form

$$E(\mathcal{C}_{gi}) = \mu; \text{Var}(\mathcal{C}_{gi}) = D; \text{Cov}(\mathcal{C}_{gi}, \mathcal{C}_{gj}) = C; \text{Cov}(\mathcal{C}_{gi}, \mathcal{C}_{hk}) = B \quad (33)$$

where the the vector μ and matrices D , C , and B do not depend upon either the clusters or the individuals. Notice that such a scenario removes the explicit dependence upon the cluster population size N in the judgement of exchangeability of the cluster population means. Suppose that we proceed under the model given by (33). In this case, Lemma 2 shows that the second stage of the adjustment of $\langle \mu_{(M;N)}(\mathcal{C}) \rangle$ given $\mathcal{C}(m; n)$ requires the adjustment of $\langle \mu_{(m;N)}(\mathcal{C}) \rangle$ given $\mu_{(m;n)}(\mathcal{C})$. We now show that this adjustment can be viewed as sampling n of N possible individuals in an exchangeable population.

Consider the set of possible observations in the sampled clusters, $\cup_{g=1}^m \cup_{i=1}^N \mathcal{C}_{gi}$ and the collection of sampled observations $\mathcal{C}(m; n) = \cup_{g=1}^m \cup_{i=1}^n \mathcal{C}_{gi}$. Suppose that we collect

the sampled observations into n mutually exclusive sets, each one containing a single observation from each of the m sampled clusters and that we also form $N - n$ similar sets for the unsampled observations, $\cup_{g=1}^m \cup_{i=n+1}^N \mathcal{C}_{gi}$. For ease of notation, assume that we form the collection $\mathcal{C}_i(m) = \{\mathcal{C}_{1i}, \dots, \mathcal{C}_{mi}\}$ for each $i = 1, \dots, N$ and the corresponding means $\mu_m(\mathcal{C}_i) = \frac{1}{m} \sum_{g=1}^m \mathcal{C}_{gi}$. Hence, $\mathcal{C}(m; n) = \cup_{i=1}^n \mathcal{C}_i(m)$. Under the model given by (33) we have, for all $i \neq j$,

$$\begin{aligned} E(\mu_m(\mathcal{C}_i)) &= \mu; \\ \text{Var}(\mu_m(\mathcal{C}_i)) &= B + \frac{1}{m}(D - B); \end{aligned} \quad (34)$$

$$\text{Cov}(\mu_m(\mathcal{C}_i), \mu_m(\mathcal{C}_j)) = B + \frac{1}{m}(C - B). \quad (35)$$

Thus the collection $\mu_m(\mathcal{C}(N)) = \{\mu_m(\mathcal{C}_1), \dots, \mu_m(\mathcal{C}_N)\}$ is a second-order exchangeable population of size N and so, using the second-order representation theorem, we may write

$$\mu_m(\mathcal{C}_i) = \mu_{(m;N)}(\mathcal{C}) + R_{(m;N),i}(\mathcal{C})$$

where, for each $i = 1, \dots, N$, $R_{(m;N),i}(\mathcal{C})$ is uncorrelated with $\mu_{(m;N)}(\mathcal{C})$. We propose to observe n of the N members of $\mu_m(\mathcal{C}(N))$, that is we observe $\mu_m(\mathcal{C}(n)) = \{\mu_m(\mathcal{C}_1), \dots, \mu_m(\mathcal{C}_n)\}$. The sample mean $\mu_{(m;n)}(\mathcal{C}) = \frac{1}{n} \sum_{i=1}^n \mu_m(\mathcal{C}_i)$ is Bayes linear sufficient for $\mu_m(\mathcal{C}(n))$ for adjusting $\langle \mu_{(m;N)}(\mathcal{C}) \rangle$ (see Theorem 2 of Goldstein and Wooff (1998)).

Definition 5 *The underlying canonical sample directions are defined as the columns of the matrix $V = [V_1 \dots V_{v_0}]$ where $V_s = TG_s$ and G_s is the s th column of the matrix G solving the generalised eigenvalue problem*

$$TG = (T + T_0)G\Xi$$

where $\Xi = \text{diag}(\xi_1, \dots, \xi_{v_0})$ is the matrix of eigenvalues and, for $i \neq j$, $T^{-1} = \text{Var}(\mu_m(\mathcal{C}_i)) - \text{Cov}(\mu_m(\mathcal{C}_i), \mu_m(\mathcal{C}_j))$ and $T_0^{-1} = \text{Cov}(\mu_m(\mathcal{C}_i), \mu_m(\mathcal{C}_j))$. G is normed so that $G^T T G = I$ and $G^T (T + T_0) \Xi = I$. The ordered eigenvalues $1 > \xi_1 \geq \dots \geq \xi_{v_0} > 0$ are termed the underlying canonical sample resolutions.

Let V_{vs} denote the v th component of the s th canonical sample direction and, for each $s = 1, \dots, v_0$, define $Y_s \in \langle \mathcal{C} \rangle$ to be $Y_s = \sum_{v=1}^{v_0} V_{vs} X_v$. The corresponding quantities in $\langle \mu_{(m;N)}(\mathcal{C}) \rangle$, $\langle \mu_{(m;n)}(\mathcal{C}) \rangle$ and each $\langle \mu_m(\mathcal{C}_i) \rangle$ which share the same coordinate representation as Y_s are defined to be $\mu_{(m;N)}(Y_s)$, $\mu_{(m;n)}(Y_s)$ and $\mu_m(Y_{is})$ respectively. An application of Theorem 2 gives the following corollary.

Corollary 4 *For a sample of size n drawn from a population of size N , the collection $\mu_{(m;N)}(\mathcal{Y}) = \{\mu_{(m;N)}(Y_1), \dots, \mu_{(m;N)}(Y_{v_0})\}$ forms a basis for $\langle \mu_{(m;N)}(\mathcal{C}) \rangle$. The $\mu_{(m;N)}(Y_s)$ are a priori uncorrelated, and, for all samples of any size, a posteriori uncorrelated. The posterior adjusted expectation, $\mu_{[N;n]s} = E_{\mu_{(m;n)}(\mathcal{C})}(\mu_{(m;N)}(Y_s))$, and posterior adjusted precision, $r_{[N;n]s} = \text{Var}_{\mu_{(m;n)}(\mathcal{C})}^{-1}(\mu_{(m;N)}(Y_s))$, for $\mu_{(m;N)}(Y_s)$ are given by*

$$\mu_{[N;n]s} = \frac{r_{[N;0]s} \mu_{[N;0]s} + g(N, n) n r_{[N]s} \mu_{(m;n)}(Y_s)}{r_{[N;0]s} + g(N, n) n r_{[N]s}}; \quad (36)$$

$$r_{[N;n]s} = r_{[N;0]s} + g(N, n) n r_{[N]s}, \quad (37)$$

where $\mu_{[N;0]s}$, $r_{[N;0]s}$ are the prior expectation and precision for $\mu_{(m;N)}(Y_s)$, $r_{[N]s}$ is the adjusted precision for $\mu_m(Y_{is})$ given $\mu_{(m;N)}(\mathcal{C})$ and $g(N, n)$ the fpc.

Corollary 4 shows that, for each m , the canonical sample directions and resolutions, for any choice of n and N , completely summarise the adjustment of $\langle \mu_{(m;N)}(\mathcal{C}) \rangle$ given $\mu_{(m;n)}(\mathcal{C})$ and, from Lemma 1 and Lemma 2, give the second stage of the adjustment of $\langle \mu_{(M;N)}(\mathcal{C}) \rangle$ given $\mathcal{C}(m;n)$.

5.3 Full adjustment

The results of Sections 5.1 and 5.2 enable us to obtain the adjustment of any quantity in $\langle \mu_{(M;N)}(\mathcal{C}) \rangle$ given $\mathcal{C}(m;n)$ under the model given by (33). For a general $Z \in \langle \mathcal{C} \rangle$, as $\mu_{(M;N)}(\mathcal{W})$ forms a basis for $\langle \mu_{(M;N)}(\mathcal{C}) \rangle$, we have

$$\begin{aligned}\mu_{(M;N)}(Z) &= \sum_{s=1}^{v_0} \text{Cov}\{\mu_{(M;N)}(Z), \mu_{(M;N)}(W_s)\} r_{[M;0]s} \mu_{(M;N)}(W_s) \\ &= \sum_{s=1}^{v_0} \delta_s \mu_{(M;N)}(W_s).\end{aligned}$$

It follows immediately from Corollary 3 that

$$\begin{aligned}E_{\mu_{(m;N)}(\mathcal{C})}(\mu_{(M;N)}(Z)) &= \sum_{s=1}^{v_0} \delta_s \mu_{(M;m)s} \\ &= \sum_{s=1}^{v_0} \delta_s (1 - \epsilon_s) \mu_{[M;0]s} + \sum_{s=1}^{v_0} \delta_s \epsilon_s \mu_{(m;N)}(W_s); \quad (38)\end{aligned}$$

$$\text{Var}_{\mu_{(m;N)}(\mathcal{C})}(\mu_{(M;N)}(Z)) = \sum_{s=1}^{v_0} \delta_s^2 r_{[M;m]s}^{-1} \quad (39)$$

where $\epsilon_s = 1 - r_{[M;0]s} r_{[M;m]s}^{-1}$. Equations (38) and (39) give, respectively, the adjusted mean and variance for a general quantity in $\langle \mu_{(M;N)}(\mathcal{C}) \rangle$ given $\mu_{(m;N)}(\mathcal{C})$ and form the first stage of the adjustment of $\langle \mu_{(M;N)}(\mathcal{C}) \rangle$ given $\mathcal{C}(m;n)$. We now complete the adjustment using the results of Corollary 4. As $\mu_{(m;N)}(\mathcal{Y})$ forms a basis for $\langle \mu_{(m;N)}(\mathcal{C}) \rangle$ then for each $s = 1, \dots, v_0$

$$\begin{aligned}\mu_{(m;N)}(W_s) &= \sum_{t=1}^{v_0} \text{Cov}\{\mu_{(m;N)}(W_s), \mu_{(m;N)}(Y_t)\} r_{[N;0]t} \mu_{(m;N)}(Y_t) \\ &= \sum_{t=1}^{v_0} \gamma_{st} \mu_{(m;N)}(Y_t).\end{aligned} \quad (40)$$

From Lemmas 1 and 2, by substituting (40) into (38) and using (36), we have

$$\begin{aligned}E_{\mathcal{C}(m;n)}(\mu_{(M;N)}(Z)) &= E_{\mu_{(m;n)}(\mathcal{C})}\{E_{\mu_{(m;N)}(\mathcal{C})}(\mu_{(M;N)}(Z))\} \\ &= \sum_{s=1}^{v_0} \delta_s (1 - \epsilon_s) \mu_{[M;0]s} + \sum_{t=1}^{v_0} \left(\sum_{s=1}^{v_0} \delta_s \epsilon_s \gamma_{st} \right) \mu_{[N;n]t} \\ &= \sum_{s=1}^{v_0} \delta_s (1 - \epsilon_s) \mu_{[M;0]s} + \sum_{t=1}^{v_0} \left(\sum_{s=1}^{v_0} \delta_s \epsilon_s \gamma_{st} \right) (1 - \eta_t) \mu_{[N;0]t} \\ &\quad + \sum_{t=1}^{v_0} \left(\sum_{s=1}^{v_0} \delta_s \epsilon_s \gamma_{st} \right) \eta_t \mu_{(m;n)}(Y_t)\end{aligned} \quad (41)$$

where $\eta_t = 1 - r_{[N;0]t} r_{[N;n]t}^{-1}$. Similarly, from Lemmas 1 and 2, equation (39) and by substituting (40) into (38) and using (37), we have

$$\begin{aligned} \text{Var}_{\mathcal{C}(m;n)}(\mu_{(M;N)}(Z)) &= \text{Var}_{\mu_{(m;N)}(\mathcal{C})}(\mu_{(M;N)}(Z)) + \text{Var}_{\mu_{(m;n)}(\mathcal{C})}\{E_{\mu_{(m;N)}(\mathcal{C})}(\mu_{(M;N)}(Z))\} \\ &= \sum_{s=1}^{v_0} \delta_s^2 r_{[M;m]s}^{-1} + \sum_{t=1}^{v_0} \left(\sum_{s=1}^{v_0} \delta_s \epsilon_s \gamma_{st} \right)^2 r_{[N;n]t}^{-1}. \end{aligned} \quad (42)$$

Equations (41) and (42) give, respectively, the adjusted expectation and variance for any quantity in $\langle \mu_{(M;N)}(\mathcal{C}) \rangle$ in terms of the $\mu_{[M;m]s}$, $r_{[M;m]s}$, $\mu_{[N;n]t}$, $r_{[N;n]t}$ for $s, t = 1, \dots, v_0$ which can be obtained through finding the underlying canonical cluster directions and resolutions, as given in Definition 4, and the underlying canonical sample directions and resolutions, as given in Definition 5. Notice how in (42) the fpc for sampling m from M attaches itself only within the $r_{[M;m]s}$ whilst the fpc for sampling n from N attaches only within the $r_{[N;n]t}$.

Corollary 5 *Suppose that the underlying canonical cluster directions and underlying canonical sample directions share, up to a constant of proportionality, the same coordinate representation. For the adjustment of $\langle \mu_{(m;n)}(\mathcal{C}) \rangle$ given $\mathcal{C}(m;n)$ the collection $\mu_{(M;N)}(\mathcal{W})$, as defined by Corollary 3, form a basis for $\langle \mu_{(m;n)}(\mathcal{C}) \rangle$ and are a priori uncorrelated, and, for all choices of m and n , a posteriori uncorrelated. The posterior adjusted expectation, $\mu_{[M,N;m,n]s} = E_{\mathcal{C}(m;n)}(\mu_{(M;N)}(W_s))$, and posterior adjusted precision, $r_{[M,N;m,n]s} = \text{Var}_{\mathcal{C}(m;n)}^{-1}(\mu_{(M;N)}(W_s))$, for $\mu_{(M;N)}(W_s)$ are given by*

$$\mu_{[M,N;m,n]s} = \frac{r_{[M;0]s}}{r_{[M;m]s}} \mu_{[M;0]s} + \left(1 - \frac{r_{[M;0]s}}{r_{[M;m]s}} \right) \tilde{\mu}_{[N;n]s} \quad (43)$$

$$r_{[M,N;m,n]s}^{-1} = \frac{1}{r_{[M;m]s}} + \left(\frac{1}{r_{[M;0]s}} - \frac{1}{r_{[M;m]s}} \right) \frac{\tilde{r}_{[N;0]s}}{\tilde{r}_{[N;n]s}} \quad (44)$$

where $\mu_{[M;0]s}$, $r_{[M;0]s}$, $r_{[M;m]s}$ are as given in Corollary 3 and

$$\tilde{\mu}_{[N;n]s} = \frac{\tilde{r}_{[N;0]s} \tilde{\mu}_{[N;0]s} + g(N, n) n \tilde{r}_{[N]s} \mu_{(m;n)}(W_s)}{\tilde{r}_{[N;0]s} + g(N, n) n \tilde{r}_{[N]s}}, \quad (45)$$

$$\tilde{r}_{[N;n]s} = \tilde{r}_{[N;0]s} + g(N, n) n \tilde{r}_{[N]s}, \quad (46)$$

where $\tilde{\mu}_{[N;0]s}$, $\tilde{r}_{[N;0]s}$ are the prior expectation and precision for $\mu_{(m;n)}(W_s)$ and $\tilde{r}_{[N]s}$ is the adjusted precision for $\mu_m(W_{is})$ given $\mu_{(m;N)}(\mathcal{C})$.

Proof - For each $s = 1, \dots, v_0$ we have $U_s = \tau_s V_s$ for some constant τ_s where U_s , V_s are, respectively, the sth underlying canonical cluster direction and sth underlying canonical sample direction. Equations (45) and (46) then follow immediately from (36) and (37). Noting that, in this case, we have $\mu_{(m;N)}(W_s) = \tau_s \mu_{(m;N)}(Y_s)$ then in (40) we must have $\gamma_{ss} = \tau_s$ and $\gamma_{st} = 0$ for all $s \neq t$. Equation (43) then follows from (41) evaluated for $Z = W_s$. Similarly evaluating (42) we have $r_{[M,N;m,n]s}^{-1} = r_{[M;m]s}^{-1} + \epsilon_s^2 \tilde{r}_{[N;n]s}^{-1}$. As equation (31) is of the form $\mu_{[M;m]s} = (1 - \epsilon_s) \mu_{[M;0]s} + \epsilon_s \mu_{(m;N)}(W_s)$ then, taking variances, we have $\epsilon_s^2 = \text{Var}(\mu_{[M;m]s}) \tilde{r}_{[N;0]s}$. $\text{Var}(\mu_{[M;m]s})$ is the variance of $\mu_{(M;N)}(W_s)$ resolved by $\mu_{(m;N)}(\mathcal{C})$, see Goldstein and Wooff (2007; p57). Equation (44) thus follows by noting that, by definition, $\text{Var}(\mu_{[M;m]s}) = r_{[M;0]s}^{-1} - r_{[M;m]s}^{-1}$. \square

Let $Res_{[M,N;m,n]s}$ denote the resolution of $\mu_{(M;N)}(W_s)$ given $\mathcal{C}(m, n)$. Rearranging equation (44) we have

$$Res_{[M,N;m,n]s} = \left(1 - \frac{r_{[M;0]s}}{r_{[M;m]s}}\right) \left(1 - \frac{\tilde{r}_{[N;0]s}}{\tilde{r}_{[N;n]s}}\right) = Res_{[M;m]s} Res_{[N;n]s} \quad (47)$$

where $Res_{[M;m]s}$ denotes the resolution of $\mu_{(M;N)}(W_s)$ given $\mu_{(m;N)}(\mathcal{C})$ and $Res_{[N;n]s}$ the resolution of $\mu_{(M;N)}(W_s)$ given $\mu_{(m;n)}(\mathcal{C})$. Equation (47) shows how the full sampling can be viewed to act multiplicatively: the resolution for $\mu_{(M;N)}(W_s)$ given $\mathcal{C}(m, n)$ is obtained by multiplying the corresponding resolutions from the m from M and n from N sampling problems together. $Res_{[M,N;m,n]s}$ may be directly obtained from the underlying canonical cluster resolutions and underlying canonical sample resolutions. Using (16) we have

$$Res_{[M,N;m,n]s} = \frac{m\{(M-1)\psi_s + 1\}}{M\{(m-1)\psi_s + 1\}} \frac{n\{(N-1)\xi_s + 1\}}{N\{(n-1)\xi_s + 1\}}. \quad (48)$$

We expect to learn most, in terms of variance reduction, about those quantities with large correlations with the early $\mu_{(M;N)}(W_s)$. As we shall see in §6.2, the simplicity of (48) makes selection of suitable sample sizes m and n to achieve specified variance reductions over elements in $\langle \mu_{(M;N)}(\mathcal{C}) \rangle$ a straightforward task. We now illustrate the theory with an example which also demonstrates that it is not unreasonable that the underlying canonical cluster directions and underlying canonical sample directions share, up to a constant of proportionality, the same co-ordinate representation.

6 Example: summarising exchangeable observations in cluster sampling

6.1 General framework

In line with the approach taken in §4.1 we shall make a further assumption.

Assumption 5 *For each individual we judge that the v_0 measurements $\mathcal{C}_{gi} = \{X_{g1i}, \dots, X_{gv_0i}\}$ under the model given by (33) are second-order exchangeable and co-exchangeable across individuals. In this case, the model for all $g \neq h$, $i \neq j$, k , reduces to*

$$\begin{aligned} E(\mathcal{C}_{gi}) &= \mu_0 \mathbf{1}_{v_0}; \quad Var(\mathcal{C}_{gi}) = (d_1 - d_2)I_{v_0} + d_2 J_{v_0}; \\ Cov(\mathcal{C}_{gi}, \mathcal{C}_{gj}) &= (c_1 - c_2)I_{v_0} + c_2 J_{v_0}; \quad Cov(\mathcal{C}_{gi}, \mathcal{C}_{hk}) = (b_1 - b_2)I_{v_0} + b_2 J_{v_0} \end{aligned}$$

where μ_0 , d_1 , d_2 , c_1 , c_2 , b_1 , and b_2 are constants, $\mathbf{1}_{v_0}$ is the $v_0 \times 1$ vector of 1s, I_{v_0} the $v_0 \times v_0$ identity matrix, and $J_{v_0} = \mathbf{1}_{v_0} \mathbf{1}_{v_0}^T$.

The effect of this simplification is that the matrices S and S_0 in Definition 4 and matrices T and T_0 in Definition 5 have a particularly simple form: they are all of the form $aI_{v_0} + bJ_{v_0}$ for suitable constants a and b . This enables the underlying canonical cluster directions and underlying canonical sample directions to both be explicitly derived and, in this case, to share up to a constant of proportionality, the same co-ordinate representation. Thus, the results of §5 reduce to those of Corollary 5.

We first solve the m from M problem. For the solution of $SF = (S + S_0)F\Psi$ we have $\Psi = diag(\psi_1, \psi_2, \dots, \psi_{v_0})$ where

$$\psi_1 = \frac{b_1 + (v_0 - 1)b_2}{a_1 + (v_0 - 1)a_2}; \quad \psi_2 = \dots = \psi_{v_0} = \frac{b_1 - b_2}{a_1 - a_2} \quad (49)$$

with $a_1 = c_1 + \frac{1}{N}(d_1 - c_1)$ and $a_2 = c_2 + \frac{1}{N}(d_2 - c_2)$. Note that $\psi_1 \geq \psi_2 \Leftrightarrow b_2 a_1 \geq a_2 b_1$ so that the ordering of the eigenvalues depends upon the prior choice. In a similar fashion to §4.1, and normed as in Definition 4, we take the matrix $F = H_{v_0} \tilde{\Psi}$ where

$$\tilde{\Psi} = \text{diag}(\sqrt{\{a_1 + (v_0 - 1)a_2\}(1 - \psi_1)}, \sqrt{(a_1 - a_2)(1 - \psi_2)}, \dots, \sqrt{(a_1 - a_2)(1 - \psi_2)}).$$

The underlying canonical cluster directions are the columns of the matrix $U = SF = H_{v_0} \tilde{\Psi}^{-1}$. We then form the collection $\mu_{(M:N)}(\mathcal{W}) = \{\mu_{(M:N)}(W_1), \dots, \mu_{(M:N)}(W_{v_0})\}$ where

$$\mu_{(M:N)}(W_1) = \frac{1}{\sqrt{v_0} \tilde{\Psi}_{11}} \sum_{v=1}^{v_0} \mu_{(M:N)}(X_v); \quad (50)$$

$$\mu_{(M:N)}(W_s) = \frac{1}{\sqrt{s(s-1)} \tilde{\Psi}_{22}} \left\{ (s-1) \mu_{(M:N)}(X_s) - \sum_{v=1}^{s-1} \mu_{(M:N)}(X_v) \right\}, \quad (51)$$

for $s = 2, \dots, v_0$. The $\mu_{(M:N)}(W_s)$ form an orthogonal grid over $\langle \mu_{(M:N)}(\mathcal{C}) \rangle$ which completely summarises the adjustment of $\langle \mu_{(M:N)}(\mathcal{C}) \rangle$ given $\mu_{(m:N)}(\mathcal{C})$. The adjusted expectation and precision for each $\mu_{(M:N)}(W_s)$ are, from Corollary 3, given by (31) and (32) where $r_{[M;0]s} = \frac{1-\psi_s}{\psi_s + \frac{1}{M}(1-\psi_s)}$ and $r_{[M]s} = \frac{M}{M-1}$.

We now solve the n from N problem. Under the model given by Assumption 5 then from (34) we can write $\text{Var}(\mu_m(\mathcal{C}_i)) = (f_1 - f_2)I_{v_0} + f_2 J_{v_0}$ and, from (35), $\text{Cov}(\mu_m(\mathcal{C}_i), \mu_m(\mathcal{C}_j)) = (e_1 - e_2)I_{v_0} + e_2 J_{v_0}$. For the solution of $TG = (T + T_0)G\Xi$ we have $\Xi = \text{diag}(\xi_1, \xi_2, \dots, \xi_{v_0})$ where

$$\xi_1 = \frac{e_1 + (v_0 - 1)e_2}{f_1 + (v_0 - 1)f_2}; \xi_2 = \dots = \xi_{v_0} = \frac{e_1 - e_2}{f_1 - f_2} \quad (52)$$

and $\xi_1 \geq \xi_2 \Leftrightarrow e_2 f_1 \geq f_2 e_1$. Note that the ξ_s depend upon m in a particularly simple fashion. We take the matrix $V = H_{v_0} \tilde{\Xi}$ where

$$\tilde{\Xi} = \text{diag}(\sqrt{\{f_1 + (v_0 - 1)f_2\}(1 - \xi_1)}, \sqrt{(f_1 - f_2)(1 - \xi_2)}, \dots, \sqrt{(f_1 - f_2)(1 - \xi_2)})$$

and V is thus normed as in Definition 5. The underlying canonical sample directions are the columns of the matrix $V = TG = H_{v_0} \tilde{\Xi}^{-1}$. We then form the collection $\mu_{(m:N)}(\mathcal{Y}) = \{\mu_{(m:N)}(Y_1), \dots, \mu_{(m:N)}(Y_{v_0})\}$ where

$$\mu_{(m:N)}(Y_1) = \frac{1}{\sqrt{v_0} \tilde{\Xi}_{11}} \sum_{v=1}^{v_0} \mu_{(m:N)}(X_v); \quad (53)$$

$$\mu_{(m:N)}(Y_s) = \frac{1}{\sqrt{s(s-1)} \tilde{\Xi}_{22}} \left\{ (s-1) \mu_{(m:N)}(X_s) - \sum_{v=1}^{s-1} \mu_{(m:N)}(X_v) \right\}, \quad (54)$$

for $s = 2, \dots, v_0$. Notice that, up to scaling constants, (53) and (54) have no dependence upon m . From Corollary 4, the adjusted expectation and precision for each $\mu_{(m:N)}(Y_s)$ are given by (36) and (37) where $r_{[N;0]s} = \frac{1-\xi_s}{\xi_s + \frac{1}{N}(1-\xi_s)}$ and $r_{[N]s} = \frac{N}{N-1}$. The $\mu_{(m:N)}(Y_s)$ form an orthogonal grid over $\langle \mu_{(m:N)}(\mathcal{C}) \rangle$ which completely summarises the adjustment of $\langle \mu_{(m:N)}(\mathcal{C}) \rangle$ given $\mu_{(m;n)}(\mathcal{C})$.

For all choices of M , N , m and n , comparing equation (50) with (53) and (51) with (54), for each $s = 1, \dots, v_0$, $\mu_{(M:N)}(W_s)$ shares, up to a constant of proportionality,

the same co-ordinate representation as $\mu_{(m;N)}(Y_s)$. Consequently, from Corollary 5, the collection $\mu_{(M;N)}(\mathcal{W})$ forms an orthogonal grid over $\langle\mu_{(M;N)}(\mathcal{C})\rangle$ which, as well as the adjustment of $\langle\mu_{(M;N)}(\mathcal{C})\rangle$ given $\mu_{(m;N)}(\mathcal{C})$, completely summarises the adjustment of $\langle\mu_{(M;N)}(\mathcal{C})\rangle$ given $\mathcal{C}(m;n)$. We now illustrate how we can use these results for setting appropriate sample sizes for both m and n .

6.2 School examination test

To illustrate the theory, we consider an examination competition where 200 schools were invited to enter 150 students. In a similar spirit to §4.2, each candidate was required to answer six compulsory questions, each marked out of 10. Let X_{gvi} denote the mark on question v of the i th candidate in the g th school. As part of the marking process, the chief examiner wishes to sample schools and individuals in order to answer a number of questions of interest, such as whether or not all the questions are of approximately the same level of difficulty. The chief examiner judges that the model given by (33) is appropriate and also that the questions are second-order exchangeable so that the model reduces to that given in Assumption 5 and we can use the framework of §6.1. Inspired by the example in §5 of Goldstein (1988), he specifies $\mu_0 = 6.5$, $d_1 = 5$, $d_2 = 3.75$, $c_1 = 1$, $c_2 = 0.25$, $b_1 = 0.3$ and $b_2 = 0.05$.

The chief examiner first finds the underlying canonical cluster directions and resolutions. Using (49) he obtains $\psi_1 = \frac{165}{718}$ and $\psi_2 = \frac{75}{226}$ (notice here that $\psi_2 > \psi_1$ but, for simplicity of exposition, we retain the ordering of §6.1) and using the resolutions forms the $\mu_{200;150}(W_s)$ as given by (50) and (51) with $v_0 = 6$. He next finds the underlying canonical sample directions and resolutions. From (52) he finds $\xi_1 = \frac{34+11m}{464+11m}$ and $\xi_2 = \frac{2+m}{4+m}$ and, using the resolutions, forms the $\mu_{(m;150)}(W_s)$ as given by (53) and (54).

The chief examiner applies Corollary 5: the collection $\mu_{(200;150)}(\mathcal{W}) = \{\mu_{200;150}(W_1), \dots, \mu_{200;150}(W_6)\}$ forms an orthogonal grid over $\langle\mu_{(200;150)}(\mathcal{C})\rangle$ which completely summarises the adjustment of $\langle\mu_{(200;150)}(\mathcal{C})\rangle$ given $\mathcal{C}(m;n)$. $\mu_{(200;150)}(W_1)$ is proportional to the average overall mark on the examination which gives insight about the general standard of the examination, whilst the $\mu_{(200;150)}(W_s)$, $s = 2, \dots, 6$ can be interpreted as summarising all of the differences in difficulty of the questions. As the grid has no dependence upon the choice of m and n (and would, under the chief examiner's beliefs, remain the same for any values of M and N) it is straightforward to explore the effect of difference choices of m and n . For example, (48) can be utilised to establish the effect of various sample sizes. For example, a choice of $m = 25$ and $n = 20$ will resolve 84.6% of the prior variance of $\mu_{(200;150)}(W_1)$ and 93.2% of that for the $\mu_{(200;150)}(W_s)$, $s = 2, \dots, 6$. The shaded areas in Figure 2(a)-(c) show, respectively, choices of m and n which achieve a proportional variance reduction of at least 0.9, 0.95 and 0.99. The corresponding plots for $\mu_{(200;150)}(W_s)$, $s = 2, \dots, 6$ are given in Figure 2(d)-(f). As there are only two distinct resolutions corresponding to $\mu_{(200;150)}(\mathcal{W})$ then, for any choice of m and n , the resolutions provide upper and lower bounds on the proportional variance reduction obtained. For example, the smallest resolution corresponds to $\mu_{(200;150)}(W_1)$ so any choice of m and n in the shaded region given by Figure 2(c) will achieve a proportional variance resolution of at least 0.99 for every quantity in $\langle\mu_{(200;150)}(\mathcal{C})\rangle$. Simple modifications can be made to any analysis using (48) to incorporate situations where, for example, the chief examiner may be faced with costs of sampling and wishes to choose m and n to achieve a specified proportional variance reduction for minimal cost.

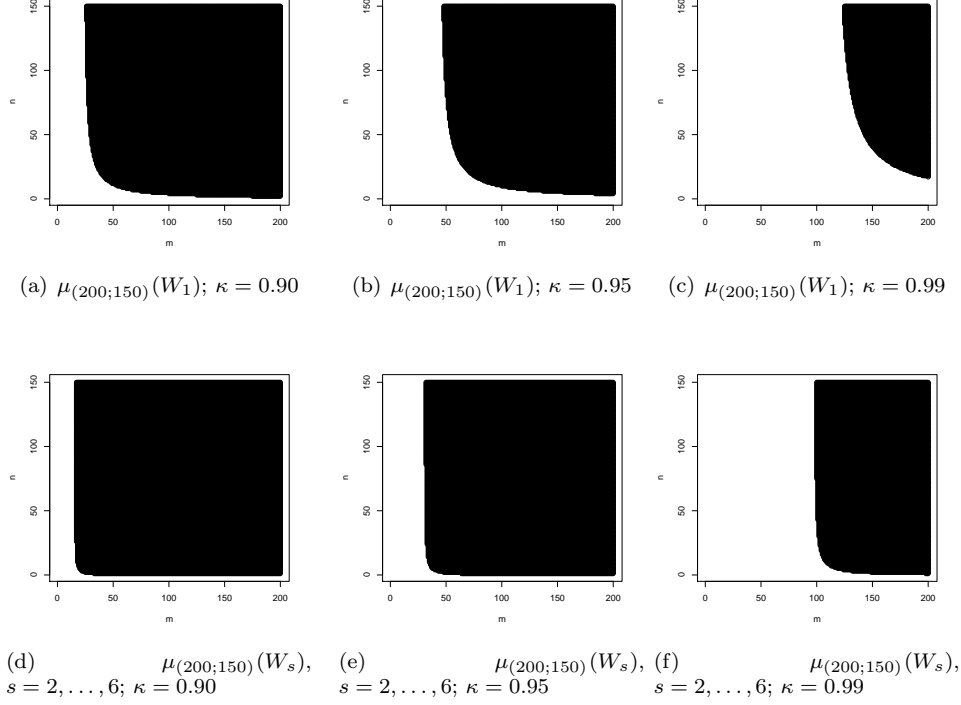


Figure 2: Shaded region shows valid choices of m and n required to achieve a proportional variance reduction of at least κ for $\mu_{(200;150)}(W_1)$ and each $\mu_{(200;150)}(W_s)$, $s = 2, \dots, 6$.

7 Conclusion

In the first half of this paper we have shown that the familiar finite population corrections naturally generalise to individuals in the multivariate population and that the types of information we gain by sampling are identified with the orthogonal canonical variable directions derived from a generalised eigenvalue problem. These ideas and techniques can also be extended to more complex systems. In the second half of the paper we explored the case of two-stage cluster sampling where each stage can be viewed as multivariate sampling from a finite population. In illustrating the theory in §4 and §6 we focused attention upon cases where additional exchangeability assumptions on the prior variable specifications meant that the generalised eigenvalue problems had simple solutions but stress that the same methods, those derived in §3 and §5, can be utilised when the prior specification does not yield such straightforward solutions to the problems given by Definitions (2), (4) and (5).

Appendix

Proof of Theorem 2 - The adjusted expectation and precision matrix for $\mu_N(\mathcal{C})$ given $\mathcal{C}(n)$ are as given by (7) and (8). We have that $\mu_N(Y_s) = H_s^T R \mu_N(\mathcal{C})$ so that, as $\text{Var}\{\mu_N(\mathcal{C})\} = R_0^{-1} + \frac{1}{N}R^{-1}$, $\text{Cov}\{\mu_N(Y_s), \mu_N(Y_t)\} = H_s^T R R_0^{-1} R H_t + \frac{1}{N}H_s^T R H_t$. From Definition 2 we have $RH(I - \Phi) = R_0 H \Phi$ with $H^T R H = I$ so that, for $s \neq t$,

$Cov\{\mu_N(Y_s), \mu_N(Y_t)\} = 0$ and

$$r_{(N;0)s}^{-1} = \frac{\phi_s}{1 - \phi_s} + \frac{1}{N} = \frac{1 + (N-1)\phi_s}{N(1 - \phi_s)}.$$

Similarly, $r_{(N)s}^{-1} = \frac{N-1}{N} H_s^T R H_s = \frac{N-1}{N}$. Using (10) we have

$$RH = (R_0 + aR)H\Phi\{I + (a-1)\Phi\}^{-1} \quad (A1)$$

Note that

$$\begin{aligned} R_{(N;0)}H &= NR_0(R_0 + NR)^{-1}RH \\ &= NR_0H\Phi\{I + (N-1)\Phi\}^{-1} \end{aligned} \quad (A2)$$

where (A2) follows from (A1) with $a = N$. Now as $R_0H\Phi = RH(I - \Phi)$ and $NR = (N-1)R_{(N)}$ then, using (A2), we have that

$$R_{(N)}H = \{R_{(N;0)} + R_{(N)}\}H\Lambda \quad (A3)$$

where $\Lambda = N^{-1}\{I + (N-1)\Phi\}$. In a similar vein to (A1), we may use (A3) to show that, for $\alpha = (1 - \frac{n-1}{N-1})^{-1}n$, we have $R_{(N)}H = \{R_{(N;0)} + \alpha R_{(N)}\}H\Lambda_{(\alpha)}$ where

$$\Lambda_{(\alpha)} = \frac{N-n}{N(N-1)}\{I + (N-1)\Phi\}\{I + (n-1)\Phi\}^{-1}.$$

Thus, $H^T R\{R_{(N;0)} + \alpha R_{(N)}\}^{-1}RH = \frac{N-1}{N}\Lambda_{(\alpha)}$. Hence, using (8), for $s \neq t$, $Cov_{\mathcal{C}(n)}\{\mu_N(Y_s), \mu_N(Y_t)\} = 0$ and

$$r_{(N;n)s}^{-1} = \frac{(N-n)\{1 + (N-1)\phi_s\}}{N^2\{1 + (n-1)\phi_s\}}. \quad (A4)$$

Now,

$$\begin{aligned} r_{(N;0)s} + \left(1 - \frac{n-1}{N-1}\right)^{-1} r_{(N)s} &= \frac{N(1 - \phi_s)}{1 + (N-1)\phi_s} + \frac{(N-1)n}{N-n} \frac{N}{N-1} \\ &= \frac{N^2\{1 + (n-1)\phi_s\}}{(N-n)\{1 + (N-1)\phi_s\}} \end{aligned}$$

which, by comparing with (A4), confirms (15). As $RH(I - \Phi) = R_0H\Phi$ then (A2) may be expressed as $R_{(N;0)}H = NRH(I - \Phi)\{I + (N-1)\Phi\}^{-1}$. Consequently, we may note that

$$\begin{aligned} H^T R\{R_{(N;0)} + \alpha R_{(N)}\}^{-1}\{R_{(N;0)}m_0 + \alpha R_{(N)}\bar{\mathcal{C}}(n)\} &= \\ \Lambda_{(\alpha)}[(N-1)(I - \Phi)\{I + (N-1)\Phi\}^{-1}H^T Rm_0 + \alpha H^T R\bar{\mathcal{C}}(n)] \end{aligned}$$

and so, from (7), we have that

$$\mu_{(N;n)s} = r_{(N;n)s}^{-1} \left\{ \frac{N(1 - \phi_s)}{1 + (N-1)\phi_s} \mu_{0s} + \alpha \frac{N}{N-1} \bar{Y}_s \right\}$$

which gives (14). \square

Proof of Lemma 1 - Note that for all g, h , $Cov(\bar{\mathcal{C}}_g, \mathcal{C}_{hj}) = Cov(\bar{\mathcal{C}}_g, \bar{\mathcal{C}}_h)$. Letting $J_{1,s}$ denote the $1 \times s$ vector of ones and \otimes the Kronecker product, we may write

$$Cov(\bar{\mathcal{C}}(m; n), \mathcal{C}(m; n)) = [J_{1,n} \otimes Var_1(\bar{\mathcal{C}}(m; n)) \dots J_{1,n} \otimes Var_m(\bar{\mathcal{C}}(m; n))]$$

where $Var_g(\bar{\mathcal{C}}(m; n)) = Cov(\bar{\mathcal{C}}(m; n), \bar{\mathcal{C}}_g)$. Denoting by I_s the $s \times s$ identity matrix and by $I_{r,s}$ the r th row of I_s then $Var^{-1}(\bar{\mathcal{C}}(m; n))Var_g(\bar{\mathcal{C}}(m; n)) = I_{g,m} \otimes I_{v_0}$ so that

$$Var^{-1}(\bar{\mathcal{C}}(m; n))Cov(\bar{\mathcal{C}}(m; n), \mathcal{C}(m; n)) = [J_{1,n} \otimes I_{1,m} \otimes I_{v_0} \dots J_{1,n} \otimes I_{m,m} \otimes I_{v_0}]. \quad (A5)$$

Now $Cov(\mu_{(M;N)}(\mathcal{C}), \bar{\mathcal{C}}_g) = Var(\mu_{(M;N)}(\mathcal{C}))$ so that

$$Cov(\mu_{(M;N)}(\mathcal{C}), \bar{\mathcal{C}}(m; n)) = J_{1,m} \otimes Var(\mu_{(M;N)}(\mathcal{C})). \quad (A6)$$

Thus, using (A5) and (A6),

$$\begin{aligned} Cov(\mu_{(M;N)}(\mathcal{C}), \bar{\mathcal{C}}(m; n))Var^{-1}(\bar{\mathcal{C}}(m; n))Cov(\bar{\mathcal{C}}(m; n), \mathcal{C}(m; n)) &= \\ &= [J_{1,n} \otimes J_{1,m}I_{1,m} \otimes Var(\mu_{(M;N)}(\mathcal{C})) \dots J_{1,n} \otimes J_{1,m}I_{1,m} \otimes Var(\mu_{(M;N)}(\mathcal{C}))] \\ &= J_{1,mn} \otimes Var(\mu_{(M;N)}(\mathcal{C})) \\ &= Cov(\mu_{(M;N)}(\mathcal{C}), \mathcal{C}(m; n)) \end{aligned} \quad (A7)$$

since $Cov(\mu_{(M;N)}(\mathcal{C}), \mathcal{C}_{hj}) = Var(\mu_{(M;N)}(\mathcal{C}))$. Now $Var(\mu_{(m;N)}(\mathcal{C})) = Cov(\mu_{(m;N)}(\mathcal{C}), \bar{\mathcal{C}}_h)$ so that $Var^{-1}(\mu_{(m;N)}(\mathcal{C}))Cov(\mu_{(m;N)}(\mathcal{C}), \bar{\mathcal{C}}(m; n)) = J_{1,m} \otimes I_{v_0}$. Further, $Cov(\mu_{(M;N)}(\mathcal{C}), \mu_{(m;N)}(\mathcal{C})) = Var(\mu_{(M;N)}(\mathcal{C}))$ so that, using (A6), we have that

$$Cov(\mu_{(M;N)}(\mathcal{C}), \mu_{(m;N)}(\mathcal{C}))Var^{-1}(\mu_{(m;N)}(\mathcal{C})), \bar{\mathcal{C}}(m; n)) = Cov(\mu_{(M;N)}(\mathcal{C}), \bar{\mathcal{C}}(m; n)). \quad (A8)$$

Results 1. and 2. follow, respectively, from (A7) and (A8) using Theorems 5.20 and 5.23 of Goldstein and Wooff (2007). \square

8 References

- Bernardo, J. M. & Smith, A.F.M. (1994). *Bayesian theory*. Wiley, Chichester.
- Bolfarine, H. (1990). Bayesian linear prediction in finite populations. *Ann. Inst. Statist. Math.* **42**, 435–444.
- Cochran, W.G. (1977). *Sampling techniques*, 3rd ed. Wiley, New York.
- De Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré* **7**, 1–68. English translation in *Studies in Subjective Probability*, Ed. H.E.Kyburg, jr. & H.E.Smokler, pp. 93–158. Wiley, New York, 1964.
- Ericson, W.A. (1969). Subjective Bayesian Models in Sampling Finite Populations (with Discussion). *J. Roy. Statist. Soc. Ser. B* **31**, 195–233.
- Ghosh, M. & Meeden, G. (1997). *Bayesian Methods for Finite Population Sampling*. Chapman & Hall, Bury St Edmunds.
- Goldstein, M. (1986). Exchangeable belief structures. *J. Amer. Statist. Assoc.* **81**, 971–6.

- Goldstein, M. (1988). The data trajectory (with discussion). In *Bayesian Statistics 3*, Ed. J. M. Bernardo, M. H. DeGroot, D. V. Lindley & A. F. M. Smith, pp. 189–209. Oxford University Press, Oxford.
- Goldstein, M. (1999). Bayes linear analysis. In *Encyclopedia of Statistical Sciences Update Volume 3*, Ed. S. Kotz, C. B. Read & D. L. Banks, pp. 29–34. Wiley, Chichester.
- Goldstein, M. & O’Hagan, A. (1996). Bayes Linear Sufficiency and Systems of Expert Posterior Assessments *J. Roy. Statist. Soc. Ser. B* **58**, 301–316.
- Goldstein, M. & Wooff, D.A. (1997). Choosing sample sizes in balanced experimental designs: a Bayes linear approach. *The Statist.* **46**, 167–183.
- Goldstein, M. & Wooff, D.A. (1998). Adjusting exchangeable beliefs. *Biometrika* **85**, 39–54.
- Goldstein, M. & Wooff, D.A. (2007). *Bayes Linear Statistics: Theory and Methods*. Wiley, Chichester.
- Hartigan, J.A. (1969). Linear Bayesian methods *J. Roy. Statist. Soc. Ser. B* **31**, 446–454.
- Hewitt, E. & Savage, L.J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80**, 470–501.
- Little, R.J.A. & Zheng, H. (2007). The Bayesian Approach to the Analysis of Finite Population Surveys. In *Bayesian Statistics 8*, Ed. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith & M. West, pp. 283–302. Oxford University Press, Oxford.
- O’Hagan, A. & Forster, J. (2004). *Kendall’s Advanced Theory of Statistics: Bayesian Inference. Vol 2B*, 2nd ed. Edward Arnold, London.
- Royall, R.M. & Pfeffermann, D. (1982) Balanced samples and robust Bayesian inference in finite population sampling *Biometrika* **69**, 401–409.
- San Martino, S., Singer, J.M. & Stanek, E.J., III. (2008). Performance of balanced two-stage empirical predictors of realized cluster latent values from finite populations: A simulation study. *Comp. Stat. Data Anal.* **52**, 2199–2217.
- Scott, A. & Smith T.M.F. (1969). Estimation in multi-stage surveys *J. Amer. Statist. Assoc.* **64**, 830–840.
- Smouse, E.P. (1984). A note on Bayesian least squares inference for finite population models. *J. Amer. Statist. Assoc.* **79**, 390–2.
- Stanek, E.J., III & Singer, J.M. (2004). Predicting random effects from finite population clustered samples with response error. *J. Amer. Statist. Assoc.* **99**, 1119–1130.